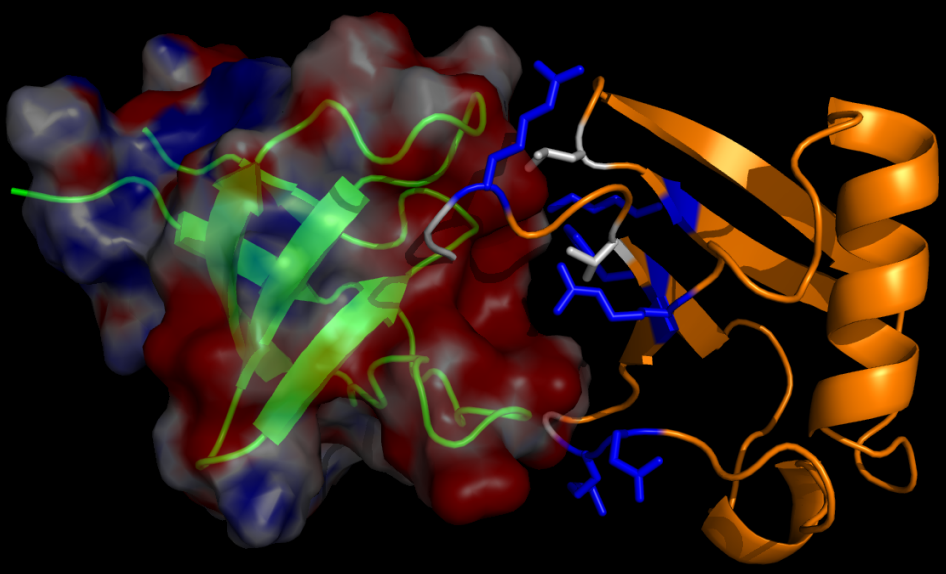
UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE QUÍMICA FÍSICA



STRUCTURE AND THERMODYNAMICS OF THE CD2AP-SH3 DOMAINS AND THEIR INTERACTION WITH UBIQUITIN

JOSÉ LUIS ORTEGA ROLDÁN

TESIS DOCTORAL

GRANADA

2010

UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE QUÍMICA FÍSICA

ESTRUCTURA Y TERMODINÁMICA DE LOS DOMINIOS SH3 DE CD2AP Y DE SU INTERACCIÓN CON UBIQUITINA

JOSÉ LUIS ORTEGA ROLDÁN

TESIS DOCTORAL

GRANADA

2010

Editor: Editorial de la Universidad de Granada
Autor: José Luis Ortega Roldán
D.L.: GR 2880-2010
ISBN: 978-84-693-2517-9

UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE QUÍMICA FÍSICA

Memoria presentada para aspirar al título de Doctor en Química (con la mención "Doctor europeo")

Fdo. **José Luis Ortega Roldán**

Licenciado en Química por la Universidad de Granada

Granada, a 11 de Marzo de 2010

VºBº

DIRECTORES DE LA TESIS

Fdo: **Nico A.J. van Nuland**
VIB Group Leader
Biomolecular NMR. Molecular Recognition Unit. Structural Biology Brussels
Vrije Universiteit Brussel/VIB

Fdo: **Ana I. Azuaga Fortes**
Profesora Ayudante Doctor
Departamento de Química Física
Universidad de Granada

A mis padres

A mi hermano

A Lorena

AGRADECIMIENTOS

En este largo periodo de algo más de cuatro años, duro en algunos momentos, pero al mismo tiempo enriquecedor, he disfrutado de las enseñanzas, ayuda y ánimo de gran cantidad de gente, a la que ahora, llegado ya el final, no deseo olvidar.

En primer lugar, deseo agradecer a Pedro que me haya aceptado dentro de su grupo de investigación, y el haber compartido su sabiduría conmigo.

Mi agradecimiento más sincero a mis jefes de tesis, Nico y Ana, por su continua dedicación y supervisión, y por haberme permitido aprender tanto en este tiempo. A Ana, por estar siempre en el despacho del fondo, dispuesta a ayudarme, y por introducirme finalmente en el mundo de la termodinámica, ¿Quién lo iba a decir?. A Nico, por haberme enseñado tanto, haberme introducido en un campo tan apasionante como el RMN, por haberme mostrado otra forma de pensar, de ver y de vivir la ciencia, por nuestros congresos, y por toda la gente que he tenido la oportunidad de conocer con él. A ambos por haberme tratado tan bien no sólo en el trabajo, sino fuera de él. Gracias por haberme permitido llegar aquí.

A Quico por estar siempre dispuesto a prestarme su ayuda, por ser mi tutor de los cursos de doctorado, además de mi "vecino" durante algunos años. A Obdulio por sus sabios consejos en las reuniones de los "CD2APs", y por ser mi tutor en la beca FPU.

A Salva, por su continuo apoyo, por su amistad, por ser mi compañero "RMNero", por introducirme en el mundo de los macs (él es el culpable), por nuestros viajes a Londres y a los "States", nuestras compras y nuestras bromas. ¡Gracias Salva!

A Irene, por iniciarme en el mundo de la investigación, A "Jose", por invitarme a entrar en este departamento, por nuestras cervezas, y por nuestras conversaciones que tanto me han ayudado. A Javi y Eva por estar siempre ahí dispuestos a ayudarme. A Enrique por nuestras charlas "futboleras", nuestros cachondeos, y por solucionar todo problema burocrático tan rápido. A María del Mar por nuestras conversaciones sobre Fórmula 1 ¡Aupa Alonso!. A José Miguel y Araceli por su gran ayuda durante este tiempo, y por nuestras conversaciones. Al resto de profesores del departamento por su apoyo.

Quiero agradecer también al Ministerio de Educación y Ciencia por la beca FPU que me ha permitido realizar esta tesis doctoral.

Al Centro de Instrumentación Científica de la Universidad de Granada, en especial a Alí, por permitirme realizar tantos experimentos de Resonancia Magnética Nuclear.

I would like to thank Martin Blackledge for giving me the opportunity of working in his Group, where I have been able to learn a lot, and at the same time I have known a group of fantastic people. I should not forget Malene for her valuable help and ideas, and Loic for our great time during our collaboration.

Special thanks to Stephanie for her support during my stay in Brussels, and for accepting to be an external referee for this thesis. Many thanks also to Nico´s group in the VUB.

No puedo olvidarme de todos mis compañeros de laboratorio, que tras nuestras largos ratos juntos, incluyendo cervezas y fiestas, han acabado convirtiéndose en mis amigos. Gracias Adela y Andrés por acogerme en el laboratorio. Por enseñarme todo lo que sabíais, por aguantarme cuando comencé. Gracias Adela por tu amistad, por ser mi compi RMNera, por nuestras fiestas en los congresos, por tus besos, por acogernos en tu casa de Londres. Sabes que puedes contar conmigo para todo. Gracias Andrés también por tu amistad, sinceridad, tu tranquilidad cuando el laboratorio lo necesitaba, y por nuestras cervezas grenoblenses. Gracias José Manuel, "Crunchito", por ser como eres. Por haber estado siempre ahí, dentro y fuera del laboratorio, por nuestras fiestas, nuestra feria y los toros ¡Viva el Fandi!, y en definitiva, por nuestra gran amistad. A M. Ángeles, "mi becaria", le agradezco que me haya aguantado todo este tiempo (incluso espadazos), que sea tan genuina, que con sus agobios sea capaz de sacarte una sonrisa en todo momento y por supuesto su amistad. A Sara, sus abrazos, besos y cariño que me han hecho sentir tan bien y que en tan poco tiempo me haya regalado su amistad. A Ana Mari, que haya sido capaz de alegrarme el día con su risa, su bondad y simpatía; por supuesto nuestras peleas vespertinas, y por haberse convertido en una gran amiga. A Javi Murciano su tranquilidad y sus conversaciones sinceras. A Manu su interés en mi trabajo. A Álvaro nuestras subidas a Sierra Nevada, a Inma su cariño, y a ambos nuestras cenas "Japo" y vuestra gran amistad. Por supuesto a mis compañeros de "zulo", Bertrand, Carles y Ángel por saber hacerme reir en cualquier momento, por su ayuda, y por haberme aguantado en los difíciles momentos de la escritura de esta Tesis. A Bertrand también por haberse interesado en todo momento por mi trabajo y por escucharme cuando me hacía falta. A Carles por nuestras conversaciones durante nuestras salidas a correr y por su ayuda en todo problema informático. A Fran, David "el Canario", María, Roci, M. Carmen, Hector, David R., y Asun, muchas gracias por nuestros agradables ratos juntos. A Mariano por nuestros desayunos juntos.

A mis amigos de la carrera, Macarena, Sisi, Luis y José Ignacio, porque junto a vosotros comencé esta andadura y por nuestros grandes momentos juntos. Especialmente le agradezco a Macarena y Manolo su gran amistad,

estemos donde estemos, y que siempre reserven un momento para pasar un rato juntos.

Por supuesto muchas gracias a mis amigos, Carlos, David, Jessy R., Jessy L, Resu, Leti, Salva y Pepe, por apoyarme tanto en estos últimos momentos, porque estar con vosotros me ha ayudado más de lo que os podéis imaginar. Porque sabéis que sois mucho más que mis amigos, y que os necesito siempre cerca de mí.

A mis padres, Carmen y Manolo, porque si he llegado hasta aquí es gracias a que hayáis decidido darlo todo por mí, porque aun sin entender muchos aspectos de mi trabajo, siempre están a mi lado para apoyarme y darme ánimos, por quererme tanto. A mi padre por ser mi modelo a seguir, porque con su trabajo me ha inculcado los valores que me han permitido llegar hasta aquí y mucho más lejos, por haber renunciado a tanto por nosotros, por nuestras largas conversaciones que tanto me han enseñado, por su infinita preocupación por mí. A mi madre por ser mi madre, valga la redundancia. Porque nunca deja de pensar en mí, por haberme enseñado tanto, por haberme guiado en mi vida, por desvivirse por mí. Sois un ejemplo para mí. Espero que os sintáis tan orgullosos de mí como yo me siento de que seáis mis padres. A mi hermano, Alberto, porque siempre he podido contar con él, porque ha estado siempre que lo necesitaba, por todo lo que hemos compartido juntos desde pequeños, porque no puede existir un hermano mejor que él. Por ser tan especial en mi vida. A M. José por su cariño, afecto y apoyo, y por saber hacerme reír tanto. A Luna, porque aun siendo tan pequeña en tan poco tiempo ha llenado mi corazón. Al resto de mi familia, mis tíos Norberto, M. Jesús, Álvaro, M. Carmen, Eva, Laura, José Miguel, Juan Manuel y M.Carmen, y primos Álvaro, Eva, Álvaro, M. Carmen, Irene, Norberto, Alejandro, Sonia, Laura, Juanma y Sergio por su apoyo incondicional. Especialmente a mi tío José Luis, porque fue el primero que me metió el gusanillo de la ciencia, porque siempre fuiste especial, y porque te echo mucho de menos. A mis abuelos, porque siempre os tengo presentes. A la familia de Lorena, especialmente a "Chico", Tere y Carlos, por acogerme como si fuera uno más y haberme dado todo su amor y cariño.

A Lorena. Porque sin ella no habría tesis. Porque no hay palabras que puedan decir todo lo que ha hecho por mí, porque sin dudarlo me ha dado todo, porque lo ha sido todo: mi compañera de laboratorio, ayudándome más que nadie, la que más me ha animado y apoyado en todo momento, la que me ha alegrado todo momento, me ha hecho todos los papeles, me ha enseñado y aconsejado tanto y tan bien, y por ser mi novia y futura esposa, porque su amor es tan grande que me hace sentirme la persona más afortunada. Por todos su cariño que me han hecho seguir adelante, por haber podido compartir todo con ella. No puedo expresar aquí todo lo que siento

por ella, pero si a alguien le debo esta tesis es a mi Lore, lo más grande que tengo en mi vida.

Y a todos aquellos que he podido olvidar, muchas gracias.

Contents

CONTENTS

Chapter 1

Introduction

The importance of protein structure

Most of the essential structure and function of cells is mediated by proteins. These macromolecules exhibit a remarkable versatility that allows them to perform a large range of activities that are fundamental to life. They are used to support the skeleton, control senses, move muscles, digest food, defend against infections and process emotions. Indeed, no other type of biological macromolecule could possibly assume all of the functions that proteins have combined over billions of years of evolution. A fundamental principle in protein science is that *protein structure leads to protein function.* The study of any protein function relies on the understanding of the protein structure, as the precise placement of particular chemical groups in specific places in the three-dimensional structure allows proteins to play important structural, transport and regulatory functions in organisms. The diversity of protein functions is translated then in a wide diversity or protein structures, from naturally unfolded proteins to large supramolecular complexes.

After completing the genome sequencing, structural proteomics projects aim a systematic mapping of the protein structural space, in order to be able to generate three dimensional structures of entire proteomes by looking just at the protein sequence. The amino acid sequence of the protein contains all the information required to acquire a proper fold, and this protein sequence is the result of the evolutionary process to achieve a reproducible, stable structure [1]. The importance of the determination of three dimensional structures of proteins relies in its intimate relationship with thermodynamics, dynamics and function.

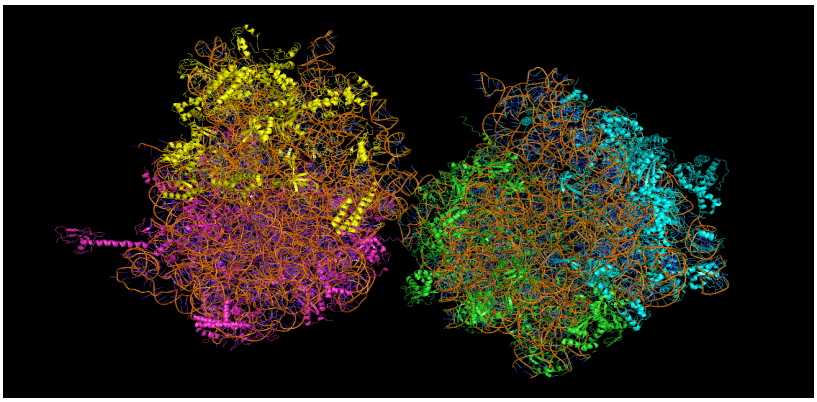


Figure 1. *Cartoon representation of the high-resolution structure of the 70S ribosome bound to the Elongation factor-P during the formation of the first peptide bond. (PDB entries 3HUW, 3HUX, 3HUY and 3HUZ).*

There are several examples on how the elucidation of high resolution structures of proteins and protein complexes provides insights into their functional mechanism. One of the most challenging structural studies addressed to date is the determination of the atomic resolution structure of the ribosome. Several structures of the ribosome bound to the mRNA and different protein factors have been determined and have revealed the mechanism by which protein synthesis takes place in the cells [2-4] (Figure 1). Another example is how the structural characterisation of the tyrosine kinase c-Abl has provided insights into the function of each domain in the full length kinase and has established the structural basis of the regulation of this complicated allosteric protein (Figure 2) [5].

Figure 2. *Cartoon representation of the high-resolution structure of full lenght c-Abl. (PDB entry 2FO0).*

Many proteins are disordered under physiological conditions, and fold into ordered structures only upon binding to their cellular targets. However, the mechanism by which this process occurs is poorly understood. Structural studies in naturally disordered proteins provide insights on how folding and binding are coupled, as for example for the interaction between the C-terminal X domain of the Measles virus phosphoprotein and the intrinsically disordered C-terminal domain of the nucleocapsid protein [6] (Figure 3).

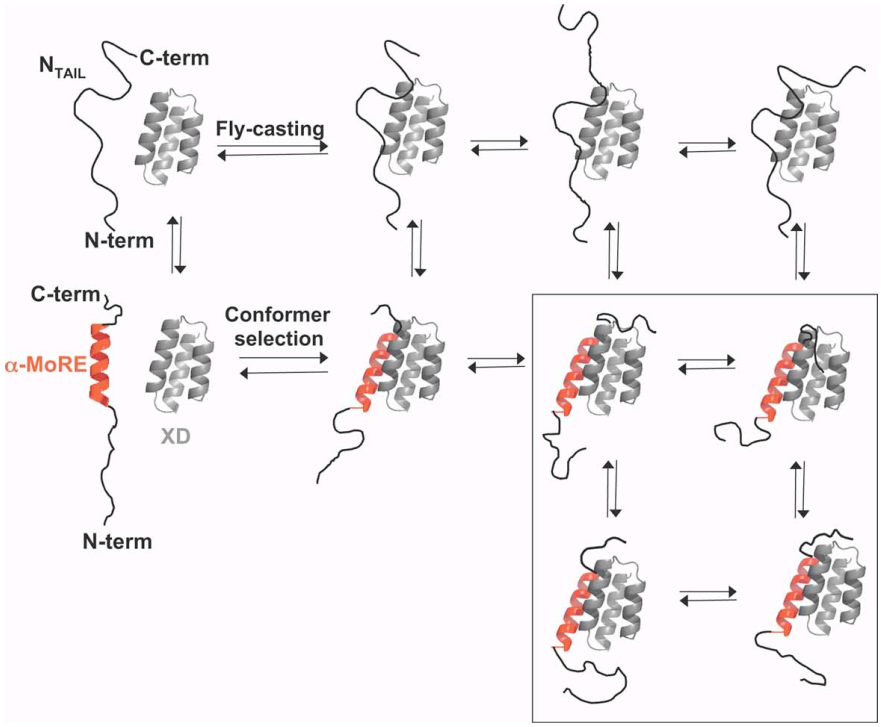


Figure 3. *Model for the mechanism of formation of the complex between the C-terminal domain Nucleocapsid-C and the terminal X domain of the Measles virus phosphoprotein. Figure taken from [6]*

Finally, an example of how the structure determination of a protein leads to the elucidation of the process of virus infection is the structural study of the mechanism of the M2 proton channel of influenza A virus [7]. In this case, this study permits to understand the effect of the anti-viral drug rimantidine and the drug-resistance mutations in the influenza A virus (Figure 4).

Three main techniques are used for the experimental structure determination of macromolecules at atomic level: X-ray crystallography, Nuclear Magnetic Resonance and Electron Microscopy. Up to date, X-ray is the most widely used, and an 86% of a total of 63559 structures determined have been solved by X-ray (Protein Data Bank statistics), while NMR has been used for the 13% of the structures deposited in the PDB and only 1% by Electron Microscopy.

X-ray crystallography is able to determine the internal structure of crystals. When an X-ray beam impinges on a crystal, a large number of scattered beams are observed in specific directions. Based on this information, the electron density distribution in the unit cell can be calculated, and then the high-resolution structure [7]. Electron microscopy, in combination with

computational three-dimensional image reconstruction, is also able to determine atomic-level structures, although its application is still minor.

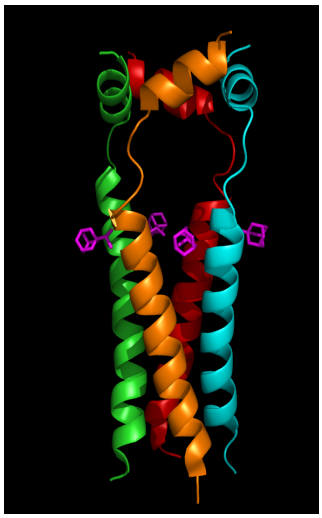


Figure 4. *Cartoon representation of the high-resolution structure of the M2 proton channel of influenza A virus bound to the viral drag rimantidine (magenta sticks). (PDB entry 2RLF).*

Nuclear Magnetic Resonance

NMR spectroscopy constitutes the second method for protein structure determination at atomic resolution in number of solved structures in the PDB. The physical processes underlying NMR and X-ray diffraction are different, and consequently the requirements for both techniques. X-ray crystallography has practically no limitation in protein size, but is restricted by the ability to grow crystals. In contrast, high resolution structure determination by NMR is currently limited to 40-50 kDa proteins. One of the largest protein structures determined by NMR using a standard NOESY approach is the 44 kDa ectodomain of simian immunodeficiency virus, (e-gp41) [8] (Figure 5). However this size limit is increasing with heteronuclear direct detection, by narrowing spectral lines at high magnetic field strengths, developing new sources of restraints for the structure calculation, as it has been shown for the structure determination of the 152 kDa transient complex between nitrite reductase and pseudoazurin [9]. On the other hand, NMR uses samples in solution, and thus NMR based data contains information relating to dynamics or macromolecular motion. This difference makes NMR spectroscopy a unique tool for the study of intrinsically disordered and partly folded proteins that are highly dynamic not amenable to X-ray crystallography, motions of molecular groups or entire domains, thermodynamic information at atomic level, etc.

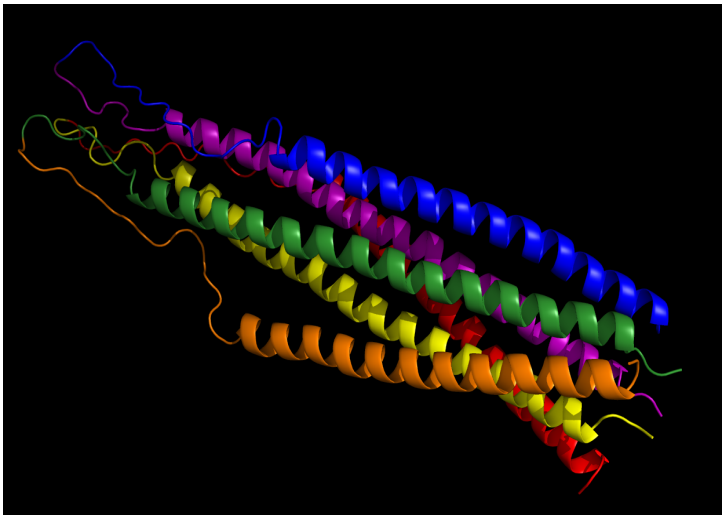


Figure 5. *Cartoon representation of the high-resolution structure of ectodomain of SIV GP41, one of the largest protein structures determined by NMR (PDB entry 2EZO).*

Protein structure determination by NMR has been traditionally based on distance restraints obtained by nuclear Overhauser enhancement spectroscopy (NOESY) [10, 11]. However, new strategies have been developed to provide either distance and geometric constraints [12], speeding up the NMR based structure determinations and allowing the study of complex systems, such as unfolded proteins, metal-binding proteins, large complexes, etc. Chemical shifts contain information from the chemical environment of the nuclei and have been exploited for the generation of structures of proteins with a maximum size of 15 kDa with reasonable accuracy, accelerating considerably the structure determination process. Recent studies [13, 14] have applied this approach for the structure determination of ubiquitin, Calbindin, HPr, Profilin, etc. Residual dipolar couplings (RDCs) and chemical shift anisotropy (CSA) constitute a source of structural information that can be exploited for ultrahigh resolution structure determination, as it has been observed for GB1 protein [15]. Additionally, RDCs permit fast structure determination, even in the absence of assigned NMR data, as it have been described for the immunoglobulin-binding domain of protein G, the C-terminal KH domain of the heterogeneous nuclear ribonucleoprotein K, ubiquitin, etc [16]. Finally, paramagnetic relaxation enhancement (PRE) and pseudoconcact shifts (PCS) have made possible to address structural studies on paramagnetic metalloproteins, such us cytochrome-c [17], that were not amenable to NMR in the past.

In addition, it has been shown recently [18] that a combination of backbone chemical shifts, residual dipolar couplings and amide proton distances from

H_N-H_N nuclear Overhauser spectroscopy permit the structure determination of large proteins up to 25 kDa without any data from the side chains using the Rosetta protein structure modelling methodology [19], thus avoiding the intensive and prone to error task of side-chain resonance assignment.

The great advantage of NMR is that, in addition to structural studies, NMR methods provide crucial information about protein dynamics and folding transitions that narrow the gap between structure and functional understanding. The ability to obtain atom-level information of dynamics is one of the hallmarks of NMR (Figure 6). There are different approaches that differ in the time scale of the motion detected [12]. ^{15}N Relaxation provides information on picosecond to nanosecond and microsecond to millisecond backbone dynamics, and therefore can reveal the functional mechanism of several proteins. For instance, such study was performed for the metallo-beta-lactamase from *Bacteroides fragilis*, where a reduction of the flexibility of its active site was found upon binding to an inhibitor, locking it and preventing substrate binding [20]. Carr-Purcell-Meiboom-Gill (CPMG)-based T_2 relaxation dispersion experiments monitor dynamics in the microsecond to millisecond time scale. It can provide information on the kinetics, thermodynamics and protein states with populations as low as 0.5 %, such as folding intermediates, as it has been described for the Fyn SH3 domain [21]. Relaxation dispersion report as well on conformational transitions in proteins by which allosteric information is transmitted and thus is also used to reveal the dynamic mechanisms underlying allosteric signal transmission [22]. Residual dipolar couplings and chemical shift anisotropy are also sensitive to protein motions [23]. Interestingly, they can detect motions from the nanosecond to millisecond time scale, constituting a unique tool for the study of dynamics in macromolecular systems, as motions in this time range cannot be detected by NMR relaxation or any other technique. One of the best examples of the use of RDCs to detect protein dynamics is the study of the structural ensemble of ubiquitin in solution. In this work it is described how ubiquitin can adopt in solution the complete structural heterogeneity found in its free and bound conformations, what is explained with conformational selection rather than induced-fit motion [24]. NMR detected Hydrogen/Deuterium (H/D) exchange permits thermodynamic studies of proteins at an atomic level. The exchange rate between the amidic protons with the D_2O present in the media detected by this type of experiments can be related to the apparent ΔG of unfolding of every residue. In this way it is possible to study the structural cooperativity of the unfolding process, as it has been observed for the α-spectrin SH3 domain [25, 26]. Finally, slower kinetic processes can be followed by direct spectral recording during the time course of the reaction in “real time” [27-29].

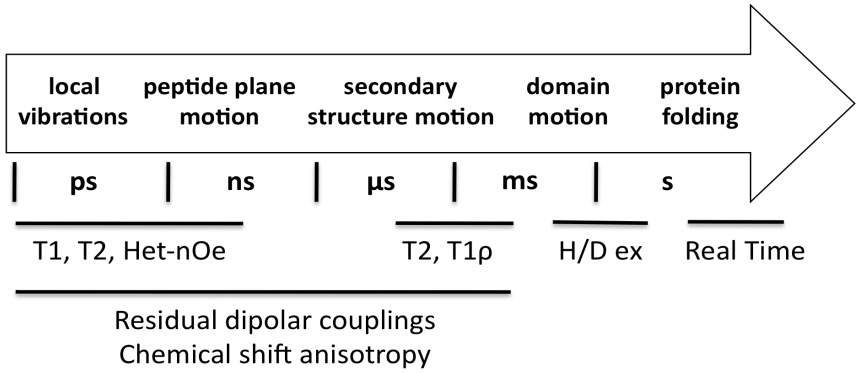


Figure 6. *Motions detected by NMR measurements.*

Thus the ability of NMR spectroscopy to address a wide variety of studies of macromolecular systems makes it the most powerful technique for a complete structural, dynamic and thermodynamic study of proteins in solution.

Biomolecular interactions

"..any phenomenon, any event, or for that matter, any 'knowledge', any transfer of information implies an interaction" [30]. Biology relies on the concerted action of a number of biomolecules organized in pathways or networks. Like in a musical composition where various notes are connected to others to give rise to a song or a supreme symphony, in the cellular interior proteins interact between each other to carry out the symphony of life. Interaction domains lead the association of proteins with one another, or with other molecules such as phospholipids, nucleic acids or small molecules, and in this sense establish a molecular language for cellular organization similar to the musical language.

Protein-protein interactions are the main responsible of the cellular machinery, which consists of millions of them acting in concert. Cellular processes require protein complexes to assemble and disassemble in very dynamic processes that have evolved over millions of years [31]. However, proteins inside the cell do not interact randomly. Protein associations need precise spatial and temporal regulation that bears remarkably functional relevance. Just single mutations or environmental factors may interfere with protein-protein interactions, leading to pathology. It is necessary then to understand the global architecture of this machinery, and how information flows through it [32].

Different types of protein-protein interactions are found depending on the type of protein chains involved, on the temporal stability of each component and on the lifetime of the complex [33]. Protein-protein interactions may occur between identical, homo-oligomeric chains, or non-identical hetero-oligomeric chains. Protein complexes can also be classified into obligate and non-obligate depending if each interacting partner is stable on its own or not. Interactions can also be distinguished based on its strength, which is described by the equilibrium dissociation constant K_d. The range of K_d values observed in biologically relevant processed is extremely wide and ranges over twelve orders of magnitude. The K_d of the binding process determines whether the interaction is weak and transient, where a continuous equilibrium of broken and formed complexes takes place, or strong and permanent, where the system only exists in its complexes form. Nevertheless, a continuum exists between the previously described types of interactions that depend on the physiological conditions and environment. Indeed, all interactions are driven by the free energy of binding and the concentration of the components. Consequently, protein-protein interactions can be regulated by influencing the binding affinity or the local concentration of each monomer. There are three types of control that take place *in vivo*: (i) Encounter of two proteins, in which co-localization is required for an efficient contact between the interacting surfaces. (ii) Control of the local concentration of each monomer. (iii) Modification of the local physicochemical environment, for example by the presence of effector molecules or changing the pH or the concentration of ions.

Nuclear Magnetic Resonance as a tool for the study of protein-protein interactions.

Up to date structural studies on protein-protein interactions have been performed mostly using X-ray crystallography, NMR spectroscopy and to a lesser extend cryo-electron microscopy. However, X-ray crystallography has the inherent limitations of crystallization of both the components of the complex in the biologically relevant conformation. Moreover, X-ray is limited to the crystallization of high-affinity complexes. NMR spectroscopy has emerged then as the most suitable technique for the structural study at atomic detail of protein-protein interactions, as it is able to characterize protein complexes under physiological conditions, even if the interactions are weak and transient.

Different strategies using NMR can be exploited, as there are a number of different classes of NMR data from which structural information can be extracted. Traditionally, NMR-based structure determination of protein complexes has relied on the detection of intermolecular NOes. Nevertheless, in the last years residual dipolar couplings (RDC), data from

paramagnetic probes (PRE), chemical shift perturbation data (CSP) and saturation transfer difference spectroscopy information (STD) have been incorporated into the structure calculation protocols of protein complexes [34].

Chemical shift perturbation analysis is the most widely used approach as a first step towards understanding protein-protein interactions. As the chemical shift is very sensitive to changes induced by the proximity of the partner protein, this experiment provides a good approximation of the binding interface. Consequently, the CSP information can be integrated in structure calculations as ambiguous interaction restraints (AIR) and are now widely used in data-drive docking calculations [35], speeding up the determination of the structure of macromolecular complexes. Several structures of protein-protein and protein-nucleic acid have been elucidated using only chemical shifts, such us the structure between barnase and two different deoxyoligonucleotides [36]. Cross-saturation also allows the detection of the binding interfaces of protein-protein interfaces with higher accuracy than CSP data [29]. Furthermore it can be used in the case of interactions with large macromolecules, as for instance membrane receptors or collagen (Figure 7) [37].

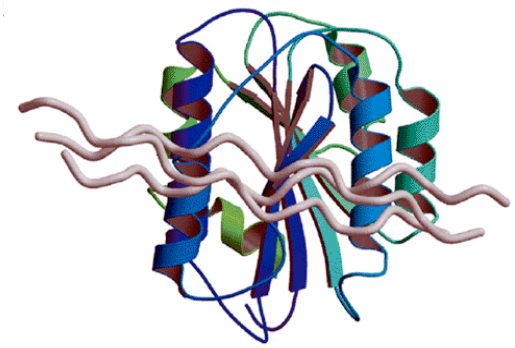


Figure 7. *Docking model of the A3 domain of von Willerbrand factor A-type and the collagen peptides [37].*

Residual dipolar couplings also provide crucial conformational constraints on intermolecular orientation in molecular complexes, or in other words, how the two partners are oriented in the complex [38]. The key long-range structural information for orienting both partners in large complexes has permitted for example structural studies on the phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr (Figure 8) [39].

Figure 8. *Structure of the Enzyme I N-terminal domain in complex with HPr (PDB entry 3EZA).*

Finally, the long-range distance restraints obtained from paramagnetic tags can be exploited for large macromolecular complex structures, as for example, the 152 kDa electron transfer complex between the copper-containing nitrite reductase (NiR) with azurin (Figure 9) [9].

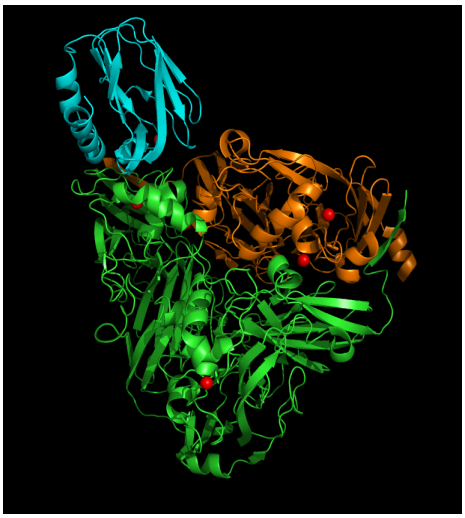


Figure 9. *Structure of the complex between azurin and NiR. The copper ions of NiR are represented as red spheres (PDB entry 2P8O).*

Adaptor proteins

As discussed before, all interactions that take place in the cellular milieu are fine regulated in time and in space. In this space, adaptor molecules take care of bringing other molecules to the correct place so that they can carry out their biological functions. Adaptor molecules are polypeptides with one or more domains able to bind other proteic or non-proteic ligands, and, in general without enzymatic activity. They have the capability to connect proteins to other proteins, to the plasma membrane or to intracellular organelles. This is how large signalling multimeric complexes are formed, localizing signalling molecules to specific subcellular localizations, and hence contributing to the specificity and efficiency of cellular responses in a spatiotemporal manner [40],[41]. The same adaptors can act at different levels in different signalling cascades. This requires lower numbers of molecules, saving the metabolic cost that their delivery would suppose to the cell. More importantly, the multi-tasking capability of adaptors allows dynamic signalling networks and integration of independent signalling events to happen through a limited number of unique proteins.

The temporal regulation is controlled by different signals both from the inside and from the outside of the cell. Activation by the epithelial growth factor (EGF) is one example of regulation triggered by an external signal. At the membrane level the EGF is recognized by the EGF receptor (EGFR) resulting in a cascade of interactions between proteins that are involved in the regulation process. The loss of this regulation and its control results in different types of diseases. Constitutive activation of EGFR is related to different types of cancer like colon, breast, bladder, cervical, ovarian, kidney etc. [42],[43],[44]. Down regulation of EGFR is hence a critical step in modulating receptor activity. It has been shown that interruption of receptor endocytosis alters duration and specificity of events that take place in this process and it has been proposed that breakdowns in this process might be as well of oncogenic character [45],[46]. Therefore, the study of the nature of events that ensure a proper modulation of duration, specificity and strength of receptor signal transduction stands out as a good target for cancer treatment.

Modular domains

To fulfil their functions, adaptor proteins use a series of different modular domains from which they are build up. Many of these domains, such as the WW, SH3, SH2 and EVH domains can be found in the human genome. Generally these domains are small (between 30 and 150 amino acids) and are folded into compact and stable structures, characterized by one or more binding sites. Each family of domains recognizes a specific and highly conserved sequence (generally between 3 to 6 amino acids) in the target

protein. In this way the WW, SH3 and EVH domains recognize proline-rich sequences, the SH2 and PTB domains peptide sequences with phosphorylated tyrosines, and the PDZ domains recognize sequences that correspond to carboxylic extremes [47],[48]. These sequences are flanked by additional fragments that interact with variable elements in the binding site and seem to determine the binding specificity within the same family of domains, in other words, how these functionally complementary proteins "*in vivo*" are related. It is interesting to point out that while other protein/protein interactions use a large contact surface, the interactions mediated by these domains imply relatively small surfaces, similar to those observed for binding sites between small ligands and enzymes. This characteristic facilitates the molecular design and production of compounds that block the protein/protein binding surfaces with high specificity [49], [50],[51]. As a consequence, these domains constitute a group of attractive targets for the design of new-generation drugs with fewer side effects.

SH3 Domains

The SH3 domains are probably the most common molecular recognition modules in the proteome [52],[53]. Until now more than 1500 different domains have been identified using database screening algorithms. They form part of proteins (such as kinases, lipases, GTPases, adaptor proteins, structural proteins and viral proteins) that participate in intracellular communication networks, in the organization of the cytoplasm skeleton and in membrane trafficking [54],[55],[56]. Frequently the SH3 domains act like an anchorage site for substrate recruitment and the formation of supra-molecular complexes that often drive the enzymatic modification of some of its components. These modifications are then translated in the production of new interaction sites and the propagation and amplification of intercellular chemical signals [57],[53]. Occasionally, these domains are essential pieces in the regulation of the enzymatic activity of the proteins they form part of by means of intra-molecular interactions with other elements of the molecule [58],[59],[60],[61]. Thus, the design of ligands that can interfere with the formation of interactions between these domains and their targets has become a viable strategy over the last years for the development of new specific drugs against different diseases [49],[46],[62],[63],[64].

All SH3 domains share the same three-dimensional structure that consists of two orthogonal sheets of three anti-parallel beta-strands (Figure 10). These strands are termed β_1, β_2, β_3, β_4, and β_5. The smaller β sheet is formed from strands β_1, part of β_2 and β_5, and the larger sheet is formed from the remainder of β_2, β_3 and β_4. Additional features of the SH3 domains are the RT (Arg-Thr) loop, the n-Src (neuronal Src) loop, the distal loop and

a 3_{10} helix. The RT loop connects strands β_1 and β_2 and has a variable sequence among SH3 domains. It contains an irregular antiparallel structure, separated by a loop containing acidic residues. The n-Src loop is formed upon insertion of 6 to 15 residues caused by a neuron-specific mRNA splicing event. It connects the strands β_2 and β_3, and as observed in the RT loop contains a variable sequence predominantly constituted of polar residues. The distal loop is a β-II type turn located between the strands β_3 and β_4, and is followed by the 3_{10} helix that separates the β_4 and β_5 strands. The effect of this particular fold is to bring most of the conserved residues close in space, resulting in the formation of an aromatic patch on the surface that forms the canonical ligand binding site for proline rich targets.

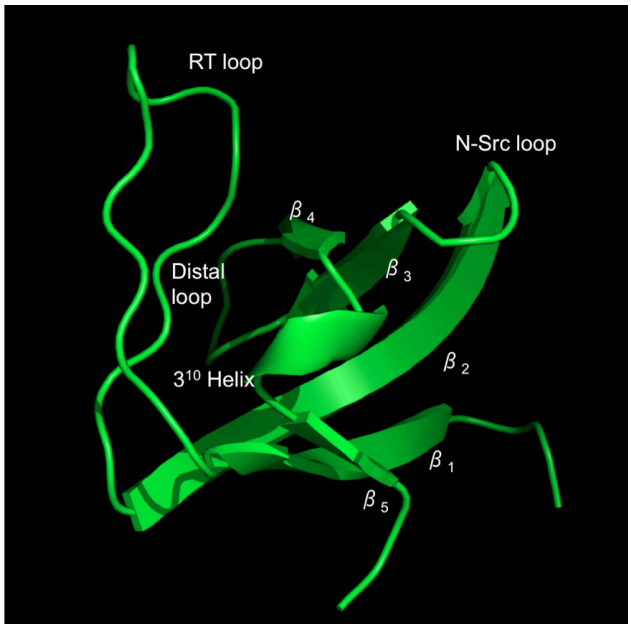


Figure 10. *Cartoon representation of the high-resolution structure of the CD2AP SH3-B domain (PDB entry 2KRN). Secondary structure elements are highlighted.*

Multiple studies on the stability of SH3 domains have been carried out so far using a wide variety of techniques, ranging from differential scanning calorimetry (DSC), circular dichroism (CD), fluorescence and stopped flow measurements. The structural stability and folding of many SH3 domains have been studied extensively by various research groups including ours [65],[66],[67],[68],[69]. Kinetic and equilibrium studies indicate that the folding-unfolding of these domains is a simple two-state process, without significantly populated intermediates. However, despite of the macroscopic simplicity of the folding/unfolding mechanism, more recent studies using NMR detected Hydrogen/Deuterium exchange have shown that the native

state of some SH3 domains needs to be considered as an ensemble of conformations whose folding reaction cannot be considered as a two-state equilibrium, as specific regions of the protein undergo independent local folding/unfolding processes. As a consequence, regional rather than the global structural cooperativity expected from a two-state equilibrium model is present in some SH3 domains [70].

Despite of the high structural homology between all the SH3 domains, differences are found in the sequence composition, allowing important differences in affinity and specificity in ligand binding. SH3 domains recognize proline-rich peptide sequences to which they generally bind with low affinity (K_d between 10^{-6} and 10^{-4} M). At the surface of the domains a relatively flat and hydrophobic surface can be identified, consisting of one acidic and two hydrophobic shallow pockets between the two β-sheets (Figure 11A). This surface is involved in the recognition and binding of proline-rich motifs [71],[72],[73]. The majority of SH3 domains bind to a proline-rich sequence with a +xφPxφP or a φPxφPx+ motif, where φ is usually a hydrophobic residue, + a basic residue (mostly Arg) and x any residue [52]. The ligand adopts an extended, left-handed helical conformation termed polyproline type II (PPII) helix. The two hydrophobic ligand-binding pockets of the SH3 domains are occupied by two hydrophobic-proline (φP) dipeptides, whereas the acidic pocket typically interacts with the basic residue in the ligand distal from the φPxφP core (Figure 11B). This basic residue binds the acidic pocket in the SH3 domain, which is thought to provide the binding specificity and denoted as the "specificity pocket". The proline-rich ligands are able to bind SH3 domains with two opposite orientations, named class I and class II. The orientation depends on the position in the sequence of the basic residue that binds the acidic pocket: ligands binding in class I orientation posses the basic residue in the N-terminus whereas class II ligands have it in the C-terminus. The difference in the location of the basic residue determines whether the N or C-terminus of the proline-rich ligand is positioned in the specificity pocket (class I and II respectively).

In spite of the fact that the shallow binding surface of an SH3 domain bears certain characteristics for the accommodation of the PPII helical structure, there are no intricate features that allow this binding surface to distinguish subtle differences between two proline-rich sequences. As a matter of fact, it has been shown that a given SH3 domain can interact with a few to several dozens of different peptide ligands [53]. It is common in SH3-interacting proteins to contain multiple weak to moderate affinity binding sites. Computational methods have indeed been developed to identify these putative canonical SH3 domain binding sequences within proline-rich regions of such proteins to narrow down the size of peptide libraries used for high throughput analysis (i-Spot, [74]; SH3-Hunter, [75]).

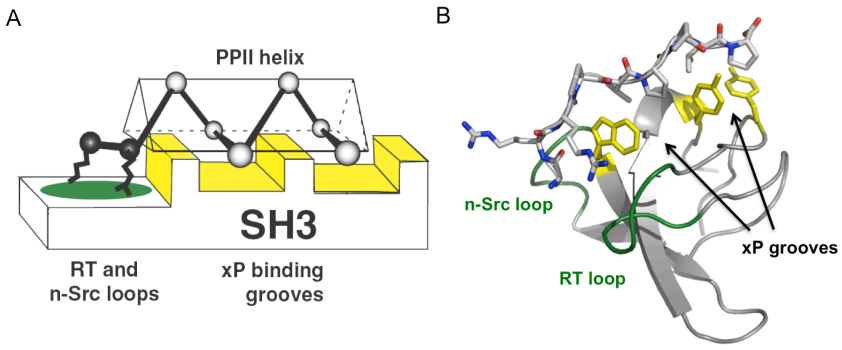


Figure 11. A. *Schematic representation of the interaction of proline-rich sequences with SH3 domains.* **B.** *Binding of the c-Src SH3 domain to a class II proline-rich ligand (Protein Data Bank code 1QWF).*

Recent studies show that certain SH3 domains are also able to bind other type of ligands, including atypical proline rich domains or even non-proline sequences. It has been described that the Eps8 L1 SH3 domain, as well as the amino-terminal SH3 domain of the adaptor protein Nck, binds a PxxDY motif from the CD2ε cytoplasmatic tail [76]. Interestingly, the SH3 domains of the adaptor proteins CD2AP and CIN85 also display an atypical polyproline recognition, binding their natural targets Cbl and CD2 through a PXXXPR motif [77]. SH3 domains are also capable of binding to non-proline-rich sequences, as observed in the interaction between the Sla1 SH3-3 [78] and CIN85 SH3 [79] domains and ubiquitin, and between the endophilin-A SH3 domain and the Parkin Ubiquitin-like domain [80].

Ubiquitin is a 76 amino-acid protein that is known to regulate a wide array of cell processes, including endocytosis, vesicular trafficking, cell-cycle control, stress response, DNA repair, signalling transcription and gene silencing via covalent ubiquitination of proteins [81]. This is a widespread regulatory post-translational covalent or non-covalent modification that can be carried out by various ubiquitin signals: monoubiquitin, multiple monoubiquitin or ubiquitin chains of diverse length and linkage. Ubiquitin binding domains (UBDs) are modular elements that can bind non-covalently to ubiquitin, and that can often distinguish different types of ubiquitin modifications. There are more than twenty families of UBDs known so far that diverge both in structure and in the type of ubiquitin recognized. Most of them use a α-helical structure to bind a hydrophobic patch in the β-sheet of ubiquitin, although there are examples where a Zinc finger or β-sheets mediate the interactions with ubiquitin [82]. SH3 domains constitute a distinct type of UBDs even though ubiquitin does not carry a proline-rich sequence. Nevertheless, ubiquitin can bind to the same hydrophobic groove on the SH3 domains competing with proline-rich sequences for this binding site [75, 76]. However, distinct binding modes

have been observed for SH3:ubiquitin interaction [83, 84], suggesting that they are involved in different biological roles.

Due to the low affinity usually exhibited for the SH3 domain interactions, the SH3 domain may dissociate quickly form one site and can subsequently be recaptured by a neighbouring site in the same molecule. Thus, the presence of multiple binding sites together with the capability of SH3 domains to recognize a collection of sequence motifs, effectively increases the local concentration of the SH3 domains and thereby promotes binding [85]. Therefore, the incidence of various SH3 binding motifs in the same molecule may be a mechanism used by the cell to regulate SH3-mediated interactions [53].

In this respect modular interaction domains often occur in tandem in regulatory proteins, affinity, and hence specificity, can often be augmented through co-operative binding involving multiple domains of the same or different kinds [86],[87]. Transient and low affinity interaction that takes place in the cell can be enhanced by the combination of more than one domain. Moreover, little is known about the effect of the presence of a second or third domain on the regulation in the affinity and specificity of the binding to a single domain; up to now most of the effort has been put on the study of isolated single domains. Multiple examples of proteins containing tandem domains exist in the genome like the CIN85/CMS (Cbl-Interacting protein of 85 kDa/Cas ligand with Multiple SH3 domains) family of adaptor proteins. In this family all three N-terminal SH3 domains are involved in a wide variety of different interactions to different targets. Moreover, these three SH3 domains share higher similarity among themselves than to any other SH3 domains, suggesting that they may have overlapping specificities in binding. It is of great interest to understand the effect of the presence of multiple SH3 domains in the mechanism of the interaction of this type of adaptor proteins.

CD2 associated protein

CD2 associated protein (CD2AP) is an adaptor protein and belongs to the known CMS/CIN85 adaptor proteins family. It was first found in T cells where CD2AP has been shown to be essential for CD2 clustering and for inducing F cell polarization at the contact area between T cell and antigen-resenting cell [88].

CD2AP is a cytosolic protein of 641 amino acids. The amino terminus consists of three SH3 domains (SH3 A, B and C) and a proline rich region, which is followed by a globular domain and a coiled-coil structure that may cause auto-dimerization of CD2AP (Figure 12). Between the second and the third SH3 domain a serine/threonine rich region is found, which serves as a

putative phosphorylation site. The proline rich region acts as a protein interaction site for multiple proteins, such as p130Cas, Fyn, Src or Yes SH3 domains. In addition, it contains four putative actin-binding sites in the C-terminus that are similar to the LKKTET motifs found in a number of actin binding proteins [89]. The three-dimensional structure of the whole protein remains unknown. However, the structure of the first, SH3-A domain was determined a couple of years ago by Bravo and co-workers using X-ray [83].

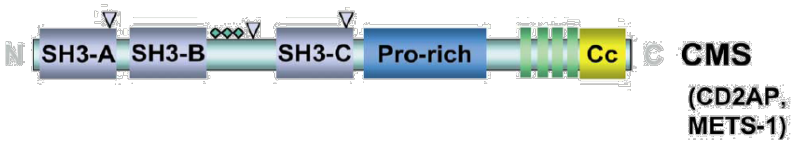


Figure 12. *The structural organization of CD2AP. It contains three N-terminal SH3 domains (Grey rectangles), a centrally located proline-rich motif (Blue rectangle), four actin-binding motifs (Green rectangles) and a C-terminal coiled-coil domain (Yellow square). Figure taken from [90].*

The renal glomerulus is a place where a variety of disorders that give rise to proteinuria and chronicle renal failure are originated [91],[92]. Recent studies on the hereditary nephritic syndrome (HNS) have emphasized the importance of the glomerular podocytes in the generation of a size-selective filtration barrier, and have provided new understanding of the mechanisms leading to proteinuria in both inherited and acquired diseases [93]. Mutations of the two genes that are implicated in this disease, encoding for the proteins nefrin, podocine and CD2AP, give rise to a severe nephritic syndrome, which suggests that these proteins are essential for an intact glomerular filter. All these proteins are found in the filtration Diaphragm or slit Diaphragm located between pedicular processes that cover the external surface of the basal glomerular membrane. This slit diaphragm is the most important structure of the glomerular capillary wall and forms the renal barrier preventing proteins to pass of the urine [94],[95],[96],[97],[98].

CD2AP knock out mice develop proteinuria, T immunodeficiency and die within the first two months of life. Rats with insufficiency in the CD2AP gene exhibit a similar phenotype to the congenital nephritic syndrome of the Finnish type (CNF). All this suggests an important role for CD2AP in the glomerular slit diaphragm, a cell-cell junction important in kidney function.

As adaptor protein, CD2AP interacts with different natural targets. CD2 is one of these targets. The interaction between the two proteins takes place via the SH3-A domain of CD2AP and a proline-rich region in the last 20 amino acids of the C-terminal part of CD2. Surface Plasma Resonance

studies have shown that this interaction is quite strong and specific [88]. In agreement with these data, CD2AP seems to cluster CD2 molecules and in this way organizing the so-called immunological Synapse. However, it is not known how CD2AP mediates the association of the CD2 molecules if only one single CD2AP molecule would interact with one CD2 molecule. A possible mechanism would involve the coil-coil region of CD2AP for the formation of homo-oligomers, or a second possible mechanism would involve not only SH3-A but also SH3-B and/or SH3-C in the interaction with CD2.

CD2AP also interacts with the product of the c-Cbl oncogene, the c-Cbl/Cbl-b protein, involved in the signal transmission network of the Tyrosinekinases, where it acts as a negative regulator [99]. The interaction with this protein and CD2AP is accomplished trough the SH3-B domain and the C-terminal part of c-Cbl which contains various proline rich motives (degeneration of the binding site) [100], although recent studies show that Cbl interaction is mediated also by the SH3-A domain [101]. This interaction site is found outside the proline-rich regions of c-Cbl where the binding to other SH3 domains takes place. Moreover, the interaction between CD2AP and c-Cbl is regulated by the phosphorylation of tyrosines in the latter protein [100]. This feature of the interaction is novel since it was thought that the binding of SH3 domains to poly-proline motifs is constituent. Interestingly, the binding of CIN85/CMS (human homologue of CD2AP) to its targets is not mediated via the classical PXXP motif but via the PXXXPR motif [102].

The Cbl family of ubiquitin ligases plays a major role in receptor ubiquitination, receptor sorting for degradation and ultimately in termination of receptor-induced signal transduction. Several lines of evidence support a role of CD2AP/CMS/CIN85 in the regulation of Cbl-directed RTK down regulation [89],[103],[104],[102]. This is a highly ordered and dynamic process controlled by the assembly of a large network of RTK-associated proteins as well as their post-translational modifications including phosphorylation and ubiquitination. As stated before, a subset of SH3 domains bind to ubiquitin [78] and this partnership is one switch in the network of interactions that control assembly and disassembly of the endocytic and actin-regulating machinery. This type of interaction is different from the PXXP type of interaction that has been related to the SH3 domains to date as being the specific binding motif. In the case of SH3:Ubiquitin binding the interaction is mediated by a complementary interface between the two proteins that can be rationalized in terms of tertiary contacts. Like shape and chemical complementarity underlies all specific molecular recognition, tertiary contacts, in a broad sense, may be a more appropriate gauge of ligand specificity for an interaction domain than in a simplified consensus motif. Domain-ligand association mediated by tertiary contacts can play important roles in signal transduction and the

SH3:Ubiquitin complex thus serves as an excellent model to study the importance of this type of interactions. Moreover, this might not only shed light on the implication of such interactions in competition with proline-rich regions of other proteins but also on their biological implications [79].

In summary we can conclude that CD2AP act as an anchoring protein that is involved in various signalling cascades controlling cellular adhesion, mobility and morphology, processes depending on the actin cytoskeleton. The diversity in type of interactions in which it is involved and the fact that CD2AP plays a role in many cellular processes makes this protein an interesting object of study from both the medical and therapeutically point of view as well as from the fundamental research point of view in order to obtain information about the mechanisms that proteins develop for the molecular recognition and self-regulation.

Aims of this thesis

Although SH3 domains have been the focus of numerous studies since their discovery about 20 years ago [105] our understanding of the mechanism of ligand recognition and signalling by this important family is far from complete. A growing body of literature suggests that SH3 may possess the most diverse specificity among interaction domains. In addition to recognizing the classical class I and II peptides that contain a PXXP core element, a number of SH3 domains have been shown to bind peptide sequences that posses alternative polyproline motifs or even lack such a motif. For example, a subset of SH3 domains that includes the SH3-3 domain of Sla1 [78], the SH3 domains of CIN85 [79] and the SH3 domains of CD2AP is able to bind ubiquitin molecules which are deficient of a proline-rich area. All this indicates that much effort is needed to obtain a complete picture of the network of interactions in which these domains are involved. In this specific case the interaction of CD2AP/CMS/CIN85 SH3-C with ubiquitin and its implications in the down regulation of the EGF asks for special attention as this process is related to a multitude of different types of cancer as mentioned earlier. A detailed structural and thermodynamic characterization of the three SH3 domains of CD2AP, as well as their interactions might give us a better understanding of not only the events in which CD2AP is involved but likely can be extended to the understanding of other similar processes within the cell.

As a first step towards understanding the biological role of the three SH3 domains of CD2AP, a detailed characterization of the structure, dynamics and thermodynamics of each domain is required. This information set the stage for further studies on their interactions with their natural targets, as well as establishes a direct connection between the structural and biophysical properties of each domain with its function inside the full-

length properties. This study is accomplished in chapters 2 and 3 of this thesis.

Because interaction domains play a central role in the assembly of multiprotein complexes, their specificity and affinity for a particular target or targets often determine the efficiency and outcome of cellular signalling [106]. It is necessary to understand the type of (weak) interaction that takes place in the endocytic pathway mediated by the SH3 domains of CD2AP. However, structural studies of weak complexes are compromised by serious practical limitations. In order to obtain high resolution structures of low affinity complexes like the structure between SH3-C and ubiquitin, novel methodologies are required. Chapter 4 of this thesis describes a new approach for structural characterization of weak macromolecular interactions by extrapolation of fully bound RDCs measured in samples with different SH3-C:Ubiquitin ratios.

Ubiquitin interaction with SH3 domains has been described for the Sla1 SH3-3 domain and the CIN85 SH3 domains and the molecular basis of this interaction were established. Interestingly, the three SH3 domains of CD2AP are able to bind ubiquitin. However, different residues were affected in SH3-C upon ubiquitin binding, suggesting a different ubiquitin binding mode. A structural characterisation of the interaction of the three SH3 domains of CD2AP with ubiquitin is needed in order to relate the different ubiquitin binding modes with their sequence determinants and biological function. Such study is performed in chapter 5 of this thesis.

Finally, all the research performed in the three SH3 domains of CD2AP during this thesis, with special emphasis on the structural aspects that drive their biological function, either in their free form or in complex with ubiquitin, is discussed in chapter 6.

References

1. Richardson, J.S., et al., *Looking at proteins: representations, folding, packing, and design. Biophysical Society National Lecture, 1992.* Biophys J, 1992. **63**(5): p. 1185-209.
2. Ban, N., et al., *Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit.* Nature, 1999. **400**(6747): p. 841-7.
3. Blaha, G., R.E. Stanley, and T.A. Steitz, *Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome.* Science, 2009. **325**(5943): p. 966-70.
4. Ban, N., et al., *The complete atomic structure of the large ribosomal subunit at 2.4 A resolution.* Science, 2000. **289**(5481): p. 905-20.

5. Nagar, B., et al., *Organization of the SH3-SH2 Unit in Active and Inactive Forms of the c-Abl Tyrosine Kinase.* Molecular Cell, 2006. **21**(6): p. 787-798.
6. Gely, S., et al., *Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein.* J Mol Recognit.
7. Drenth, J., *X-ray Diffraction: Principles.* Encyclopedia of Life Sciences, 2003.
8. Caffrey, M., et al., *Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41.* EMBO J, 1998. **17**(16): p. 4572-84.
9. Vlasie, M.D., et al., *Conformation of Pseudoazurin in the 152†kDa Electron Transfer Complex with Nitrite Reductase Determined by Paramagnetic NMR.* Journal of Molecular Biology, 2008. **375**(5): p. 1405-1415.
10. Wuthrich, K., *Protein structure determination in solution by nuclear magnetic resonance spectroscopy.* Science, 1989. **243**(4887): p. 45-50.
11. Wuthrich, K., *NMR of proteins and nucleic acids*, ed. J.W. Sons. 1986.
12. Grzesiek, S. and H.J. Sass, *From biomolecular structure to functional understanding: new NMR developments narrow the gap.* Curr Opin Struct Biol, 2009. **19**(5): p. 585-95.
13. Shen, Y., et al., *Consistent blind protein structure generation from NMR chemical shift data.* Proc Natl Acad Sci U S A, 2008. **105**(12): p. 4685-90.
14. Cavalli, A., et al., *Protein structure determination from NMR chemical shifts.* Proc Natl Acad Sci U S A, 2007. **104**(23): p. 9615-20.
15. Bouvignies, G., et al., *Ultrahigh-Resolution Backbone Structure of Perdeuterated Protein GB1 Using Residual Dipolar Couplings from Two Alignment Media13.* Angewandte Chemie International Edition, 2006. **45**(48): p. 8166-8169.
16. Meiler, J. and D. Baker, *Rapid protein fold determination using unassigned NMR data.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15404-9.
17. Bertini, I., et al., *NMR spectroscopy of paramagnetic metalloproteins.* Chembiochem, 2005. **6**(9): p. 1536-49.
18. Raman, S., et al., *NMR structure determination for larger proteins using backbone-only data.* Science. **327**(5968): p. 1014-8.
19. Das, R. and D. Baker, *Macromolecular modeling with rosetta.* Annu Rev Biochem, 2008. **77**: p. 363-82.
20. Huntley, J.J., et al., *Dynamics of the metallo-beta-lactamase from Bacteroides fragilis in the presence and absence of a tight-binding inhibitor.* Biochemistry, 2000. **39**(44): p. 13356-64.

21. Korzhnev, D.M., et al., *Abp1p and Fyn SH3 domains fold through similar low-populated intermediate states.* Biochemistry, 2006. **45**(34): p. 10175-83.
22. Bruschweiler, S., et al., *Direct Observation of the Dynamic Process Underlying Allosteric Signal Transmission.* J Am Chem Soc, 2009.
23. Blackledge, M., *Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2005. **46**(1): p. 23-61.
24. Lange, O.F., et al., *Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution.* Science, 2008. **320**(5882): p. 1471-5.
25. Casares, S., et al., *Cooperative propagation of local stability changes from low-stability and high-stability regions in a SH3 domain.* Proteins, 2007. **67**(3): p. 531-47.
26. Grantcharova, V.P. and D. Baker, *Folding dynamics of the src SH3 domain.* Biochemistry, 1997. **36**(50): p. 15685-92.
27. van Nuland, N.A.J., et al., *Real-Time NMR Studies of Protein Folding.* Accounts of Chemical Research, 1998. **31**(11): p. 773-780.
28. Mok, K.H. and P.J. Hore, *Photo-CIDNP NMR methods for studying protein folding.* Methods, 2004. **34**(1): p. 75-87.
29. Schanda, P., V. Forge, and B. Brutscher, *Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy.* Proc Natl Acad Sci U S A, 2007. **104**(27): p. 11257-62.
30. Monod, J., *On symmetry and function in biological systems*. Symmetry and function of biological systems at the macro-molecular level. 1969.
31. Levy, E.D. and J.B. Pereira-Leal, *Evolution and dynamics of protein interactions and networks.* Curr Opin Struct Biol, 2008. **18**(3): p. 349-57.
32. Bader, S., S. Kuhner, and A.C. Gavin, *Interaction networks for systems biology.* FEBS Lett, 2008. **582**(8): p. 1220-4.
33. Nooren, I.M. and J.M. Thornton, *Diversity of protein-protein interactions.* EMBO J, 2003. **22**(14): p. 3486-92.
34. O'Connell, M.R., R. Gamsjaeger, and J.P. Mackay, *The structural analysis of protein-protein interactions by NMR spectroscopy.* Proteomics, 2009. **9**(23): p. 5224-32.
35. Dominguez, C., R. Boelens, and A.M. Bonvin, *HADDOCK: a protein-protein docking approach based on biochemical or biophysical information.* J Am Chem Soc, 2003. **125**(7): p. 1731-7.
36. Cioffi, M., et al., *Use of quantitative (1)H NMR chemical shift changes for ligand docking into barnase.* J Biomol NMR, 2009. **43**(1): p. 11-9.

37. Nishida, N., et al., *Collagen-binding mode of vWF-A3 domain determined by a transferred cross-saturation experiment.* Nat Struct Biol, 2003. **10**(1): p. 53-8.
38. Vaynberg, J. and J. Qin, *Weak protein-protein interactions as probed by NMR spectroscopy.* Trends Biotechnol, 2006. **24**(1): p. 22-7.
39. Garrett, D.S., et al., *Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr.* Nat Struct Biol, 1999. **6**(2): p. 166-73.
40. Pawson, T., M. Raina, and P. Nash, *Interaction domains: from simple binding events to complex cellular behavior.* FEBS Lett, 2002. **513**(1): p. 2-10.
41. Csiszar, A., *Structural and functional diversity of adaptor proteins involved in tyrosine kinase signalling.* Bioessays, 2006. **28**(5): p. 465-79.
42. Salomon, D.S., et al., *Epidermal growth factor-related peptides and their receptors in human malignancies.* Crit Rev Oncol Hematol, 1995. **19**(3): p. 183-232.
43. Blume-Jensen, P. and T. Hunter, *Oncogenic kinase signalling.* Nature, 2001. **411**(6835): p. 355-65.
44. Schlessinger, J., *Ligand-induced, receptor-mediated dimerization and activation of EGF receptor.* Cell, 2002. **110**(6): p. 669-72.
45. Peschard, P. and M. Park, *Escape from Cbl-mediated downregulation: a recurrent theme for oncogenic deregulation of receptor tyrosine kinases.* Cancer Cell, 2003. **3**(6): p. 519-23.
46. Crosetto, N., R. Tikkanen, and I. Dikic, *Oncogenic breakdowns in endocytic adaptor proteins.* FEBS Lett, 2005. **579**(15): p. 3231-8.
47. Kay, B.K., M.P. Williamson, and M. Sudol, *The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains.* Faseb J, 2000. **14**(2): p. 231-41.
48. Kuriyan, J. and D. Cowburn, *Modular peptide recognition domains in eukaryotic signaling.* Annu Rev Biophys Biomol Struct, 1997. **26**: p. 259-88.
49. Kardinal, C., et al., *Cell-penetrating SH3 domain blocker peptides inhibit proliferation of primary blast cells from CML patients.* Faseb J, 2000. **14**(11): p. 1529-38.
50. Kardinal, C., et al., *Rational development of cell-penetrating high affinity SH3 domain binding peptides that selectively disrupt the signal transduction of Crk family adapters. Amgen Peptide Technology Group.* Ann N Y Acad Sci, 1999. **886**: p. 289-92.
51. Lee, K.Y., et al., *A dipalmitoyl peptide that binds SH3 domain, disturbs intracellular signal transduction, and inhibits tumor growth in vivo.* Biochem Biophys Res Commun, 2002. **296**(2): p. 434-42.

52. Mayer, B.J., *SH3 domains: complexity in moderation.* J Cell Sci, 2001. **114**(Pt 7): p. 1253-63.
53. Li, S.S., *Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction.* Biochem J, 2005. **390**(Pt 3): p. 641-53.
54. McPherson, P.S., *Regulatory role of SH3 domain-mediated protein-protein interactions in synaptic vesicle endocytosis.* Cell Signal, 1999. **11**(4): p. 229-38.
55. Skorski, T., et al., *The SH3 domain contributes to BCR/ABL-dependent leukemogenesis in vivo: role in adhesion, invasion, and homing.* Blood, 1998. **91**(2): p. 406-18.
56. Stein, R., *SH2 and SH3 domains. Unraveling signaling networks with peptide antagonists.* Methods Mol Biol, 1998. **88**: p. 187-95.
57. Kiyokawa, E., et al., *Role of Crk oncogene product in physiologic signaling.* Crit Rev Oncog, 1997. **8**(4): p. 329-42.
58. Arold, S.T., et al., *The role of the Src homology 3-Src homology 2 interface in the regulation of Src kinases.* J Biol Chem, 2001. **276**(20): p. 17199-205.
59. Barila, D. and G. Superti-Furga, *An intramolecular SH3-domain interaction regulates c-Abl activity.* Nat Genet, 1998. **18**(3): p. 280-2.
60. Brabek, J., et al., *The SH3 domain of Src can downregulate its kinase activity in the absence of the SH2 domain-pY527 interaction.* Biochem Biophys Res Commun, 2002. **296**(3): p. 664-70.
61. Brasher, B.B., S. Roumiantsev, and R.A. Van Etten, *Mutational analysis of the regulatory function of the c-Abl Src homology 3 domain.* Oncogene, 2001. **20**(53): p. 7744-52.
62. Oneyama, C., H. Nakano, and S.V. Sharma, *UCS15A, a novel small molecule, SH3 domain-mediated protein-protein interaction blocking drug.* Oncogene, 2002. **21**(13): p. 2037-50.
63. Stauffer, T.P., et al., *Inhibition of Lyn function in mast cell activation by SH3 domain binding peptides.* Biochemistry, 1997. **36**(31): p. 9388-94.
64. Vidal, M., V. Gigoux, and C. Garbay, *SH2 and SH3 domains as targets for anti-proliferative agents.* Crit Rev Oncol Hematol, 2001. **40**(2): p. 175-86.
65. Filimonov, V.V., et al., *A thermodynamic analysis of a family of small globular proteins: SH3 domains.* Biophys Chem, 1999. **77**(2-3): p. 195-208.
66. Guijarro, J.I., et al., *Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy.* J Mol Biol, 1998. **276**(3): p. 657-67.
67. Plaxco, K.W., et al., *The folding kinetics and thermodynamics of the Fyn-SH3 domain.* Biochemistry, 1998. **37**(8): p. 2529-37.
68. Vega, M.C., J.C. Martinez, and L. Serrano, *Thermodynamic and structural characterization of Asn and Ala residues in the*

disallowed II' region of the Ramachandran plot. Protein Sci, 2000. **9**(12): p. 2322-8.
69. Viguera, A.R., et al., *Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition.* Biochemistry, 1994. **33**(8): p. 2142-50.
70. Casares, S., et al., *Structural cooperativity in the SH3 domain studied by site-directed mutagenesis and amide hydrogen exchange.* FEBS Lett, 2003. **539**(1-3): p. 125-30.
71. Larson, S.M. and A.R. Davidson, *The identification of conserved interactions within the SH3 domain by alignment of sequences and structures.* Protein Sci, 2000. **9**(11): p. 2170-80.
72. Cesareni, G., et al., *Can we infer peptide recognition specificity mediated by SH3 domains?* FEBS Lett, 2002. **513**(1): p. 38-44.
73. Ferreon, J.C. and V.J. Hilser, *Thermodynamics of binding to SH3 domains: the energetic impact of polyproline II (PII) helix formation.* Biochemistry, 2004. **43**(24): p. 7787-97.
74. Brannetti, B. and M. Helmer-Citterich, *iSPOT: A web tool to infer the interaction specificity of families of protein modules.* Nucleic Acids Res, 2003. **31**(13): p. 3709-11.
75. Ferraro, E., et al., *SH3-Hunter: discovery of SH3 domain interaction sites in proteins.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W451-4.
76. Aitio, O., et al., *Structural basis of PxxDY motif recognition in SH3 binding.* J Mol Biol, 2008. **382**(1): p. 167-78.
77. Moncalian, G., et al., *Atypical polyproline recognition by the CMS N-terminal Src homology 3 domain.* J Biol Chem, 2006. **281**(50): p. 38845-53.
78. Stamenova, S.D., et al., *Ubiquitin binds to and regulates a subset of SH3 domains.* Mol Cell, 2007. **25**(2): p. 273-84.
79. Bezsonova, I., et al., *Interactions between the three CIN85 SH3 domains and ubiquitin: implications for CIN85 ubiquitination.* Biochemistry, 2008. **47**(34): p. 8937-49.
80. Trempe, J.F., et al., *SH3 domains from a subset of BAR proteins define a Ubl-binding domain and implicate parkin in synaptic ubiquitination.* Mol Cell, 2009. **36**(6): p. 1034-47.
81. Hurley, J.H., S. Lee, and G. Prag, *Ubiquitin-binding domains.* Biochem J, 2006. **399**(3): p. 361-72.
82. Dikic, I., S. Wakatsuki, and K.J. Walters, *Ubiquitin-binding domains - from structures to functions.* Nat Rev Mol Cell Biol, 2009. **10**(10): p. 659-71.
83. Korzhnev, D.M., et al., *Alternate binding modes for a ubiquitin-SH3 domain interaction studied by NMR spectroscopy.* J Mol Biol, 2009. **386**(2): p. 391-405.
84. Kang, J., et al., *Distinct interactions between ubiquitin and the SH3 domains involved in immune signaling.* Biochim Biophys Acta, 2008. **1784**(9): p. 1335-41.

85. Jia, C.Y., et al., *Novel Src homology 3 domain-binding motifs identified from proteomic screen of a Pro-rich region.* Mol Cell Proteomics, 2005. **4**(8): p. 1155-66.
86. Pawson, T. and P. Nash, *Assembly of cell regulatory systems through protein interaction domains.* Science, 2003. **300**(5618): p. 445-52.
87. Pawson, T., *Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems.* Cell, 2004. **116**(2): p. 191-203.
88. Dustin, M.L., et al., *A novel adaptor protein orchestrates receptor patterning and cytoskeletal polarity in T-cell contacts.* Cell, 1998. **94**(5): p. 667-77.
89. Kirsch, K.H., et al., *CMS: an adapter molecule involved in cytoskeletal rearrangements.* Proc Natl Acad Sci U S A, 1999. **96**(11): p. 6211-6.
90. Dikic, I., *CIN85/CMS family of adaptor molecules.* FEBS Lett, 2002. **529**(1): p. 110-5.
91. Kerjaschki, D., *Caught flat-footed: podocyte damage and the molecular bases of focal glomerulosclerosis.* J Clin Invest, 2001. **108**(11): p. 1583-7.
92. Kriz, W. and K.V. Lemley, *The role of the podocyte in glomerulosclerosis.* Curr Opin Nephrol Hypertens, 1999. **8**(4): p. 489-97.
93. Wolf, G. and R.A. Stahl, *CD2-associated protein and glomerular disease.* Lancet, 2003. **362**(9397): p. 1746-8.
94. Holthofer, H., et al., *Nephrin localizes at the podocyte filtration slit area and is characteristically spliced in the human kidney.* Am J Pathol, 1999. **155**(5): p. 1681-7.
95. Holzman, L.B., et al., *Nephrin localizes to the slit pore of the glomerular epithelial cell.* Kidney Int, 1999. **56**(4): p. 1481-91.
96. Ruotsalainen, V., et al., *Nephrin is specifically located at the slit diaphragm of glomerular podocytes.* Proc Natl Acad Sci U S A, 1999. **96**(14): p. 7962-7.
97. Schwarz K, S.M., Reiser J, Saleem MA, Faul C, Kriz W, Shaw AS, Holzman LB, Mundel P., *Podocin, a raft-associated component of the glomerular slit diaphragm, interacts with CD2AP and nephrin.* The Journal of Clinical Investigation, 2001. **108**(11): p. 1621-1629.
98. Shih, N.Y., et al., *CD2AP localizes to the slit diaphragm and binds to nephrin via a novel C-terminal domain.* Am J Pathol, 2001. **159**(6): p. 2303-8.
99. Lupher, M.L., Jr., et al., *The Cbl protooncoprotein: a negative regulator of immune receptor signal transduction.* Immunol Today, 1999. **20**(8): p. 375-82.
100. Kirsch, K.H., et al., *The adapter type protein CMS/CD2AP binds to the proto-oncogenic protein c-Cbl through a tyrosine*

phosphorylation-regulated Src homology 3 domain interaction. J Biol Chem, 2001. **276**(7): p. 4957-63.

101. Ababou, A., M. Pfuhl, and J.E. Ladbury, *Novel insights into the mechanisms of CIN85 SH3 domains binding to Cbl proteins: solution-based investigations and in vivo implications.* J Mol Biol, 2009. **387**(5): p. 1120-36.
102. Szymkiewicz, I., et al., *CIN85 participates in Cbl-b-mediated down-regulation of receptor tyrosine kinases.* J Biol Chem, 2002. **277**(42): p. 39666-72.
103. Petrelli, A., et al., *The endophilin-CIN85-Cbl complex mediates ligand-dependent downregulation of c-Met.* Nature, 2002. **416**(6877): p. 187-90.
104. Soubeyran, P., et al., *Cbl-CIN85-endophilin complex mediates ligand-induced downregulation of EGF receptors.* Nature, 2002. **416**(6877): p. 183-7.
105. Mayer, B.J., M. Hamaguchi, and H. Hanafusa, *A novel viral oncogene with structural similarity to phospholipase C.* Nature, 1988. **332**(6161): p. 272-5.
106. Nishida, M., et al., *Novel recognition mode between Vav and Grb2 SH3 domains.* Embo J, 2001. **20**(12): p. 2995-3007.

Chapter 2

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

Reprinted from Journal of Biomolecular NMR, 2007, Vol. 39, 331-336.
(Jose L. Ortega Roldan, M. Luisa Romero Romero, Ari Ora, Eiso AB, Obdulio Lopez Mayorga, Ana I. Azuaga and Nico A.J. van Nuland)

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}Ha, ^{13}Ca, ^{13}CO and ^{13}Cb 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.

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 CD2AP 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 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 4 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 hour in M9 medium at 20 ºC. Protein expression was induced with 1 mM IPTG at 20 ºC during 15-20 hours. Cells were resuspended after centrifugation at 6,197 g for 20 minutes in 50 mM phosphate, 0.3 M NaCl, pH 8.0 buffer (Column Buffer: CB) and broken in a French press. After centrifugation at 104,630 g during 30 minutes, 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 13980 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 25ºC on a Varian NMR Direct-Drive Systems 600 MHz spectrometer (^1H

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, HNCACB, 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}Ca and ^{13}Cb 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. ^{1}Ha and ^{1}Hb were assigned using the HBHA(CO)NH. The obtained ^{1}Ha-^{13}Ca and ^{1}Hb-^{13}Cb 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 ^{1}Hd and ^{1}He 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}Cb 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.

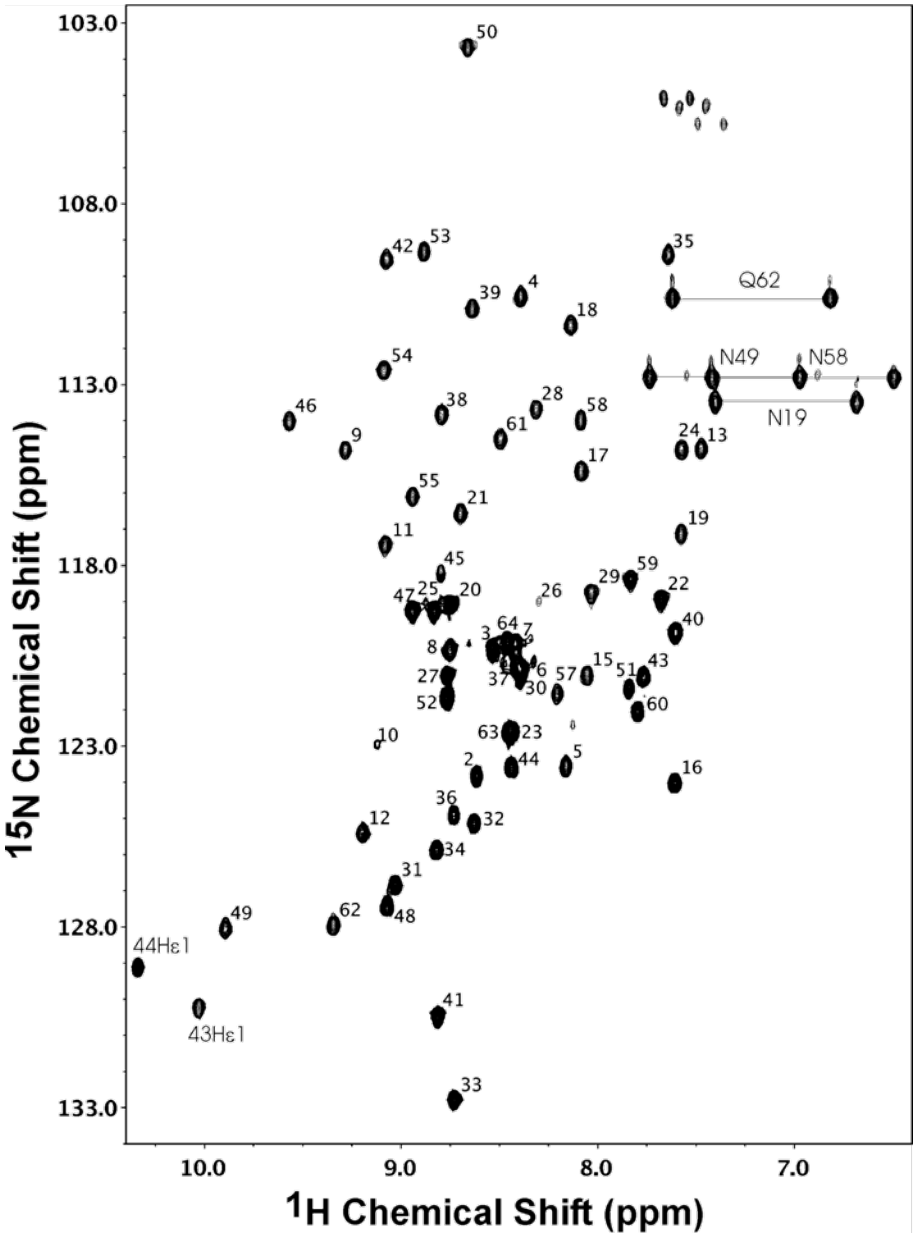


Figure 1. Assigned 600 MHz ^{1}H-^{15}N HSQC of CD2AP SH3-C at pH 2.0, 25 ºC. *All backbone ^{1}H-^{15}N pairs are indicated by their corresponding number in the amino acid sequence. The Hε1 protons of Trp43 and Trp44 as well as the side-chains of the single glutamine and three asparagines are indicated.*

The 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 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}Ha, ^{13}Ca, ^{13}CO and ^{13}Cb chemical shifts using the program PREDITOR (Berjanskii and Wishart 2006). Only *phi* and *psi* angle restraints were used with a confidence factor higher than 0.5. The standard protocol was used with 7 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 (Figure 3A).

Table 1. *Structural statistics for the 20 conformers of CD2AP SH3-C. The statistics is obtained from an ensemble of twenty lowest-energy water-refined structures for SH3-C. [2]Restraint statistics reported for unique, unambiguous assigned NOes. [3]Coordinate precision is given as the pair-wise Cartesian coordinate Root Mean Square Deviations from the average structure over the ensemble.[4]Values obtained from the PROCHECK-NMR analysis over all residues including the first seven highly flexible residues.*

Experimental restraints[2]	
Distance restraints	
Intraresidue (i-j = 0)	241
Sequential (|i-j| = 1)	297
Medium range (2£|i-j|£4)	109
Long range (|i-j|³5)	394
Dihedral restraints	
Torsion angle (phi/psi)	2x24
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)[3]	
Backbone N, CA, C'	0.37
Heavy atoms	0.91
Ramachandran plot[4]	
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

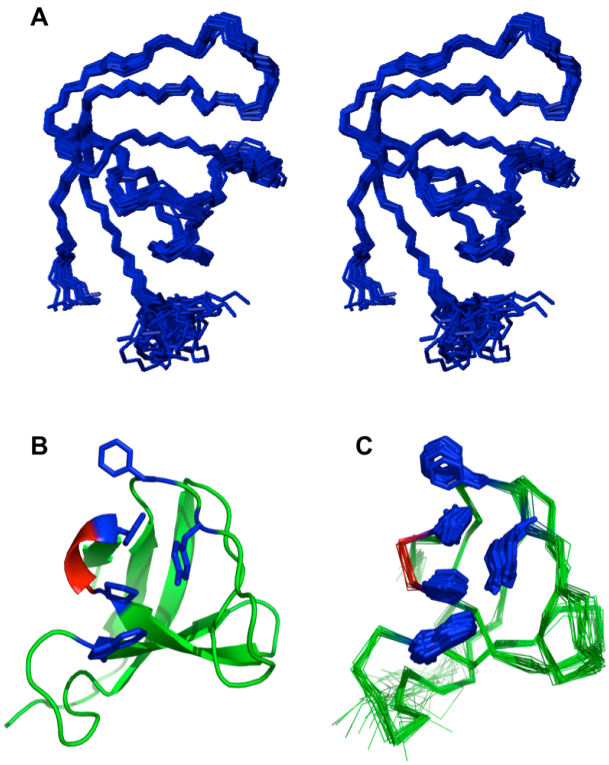


Figure 2. The high resolution NMR structure of CD2AP SH3-C. A. *Stereo picture of the ensemble of 20 lowest energy structures superimposed on the backbone atoms of residues 8 to 62.* ***B.*** *Cartoon representation of the lowest energy structure. The 5 b-strands and loops and turns are colored green. 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.*

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 ºC of CD2AP SH3-C. Relaxation values were obtained from series of 2D experiments with coherence selection achieved by pulse field gradients at 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 (Figure 3A). Analysis of the relaxation data of SH3-C using TENSOR2 (Dosset et al. 2000) indicate average ^{15}N R_2/R_1 ratios (Figure 3B) for 32 residues in the most ordered regions corresponding to an apparent rotational correlation time t_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 (Figure 3C). Besides the 7 N-terminal residues, the local internal mobility of residues 24 and 60 can only be described using an additional term for fast internal motions (t_e) or for exchange contributions (R_{ex}), respectively.

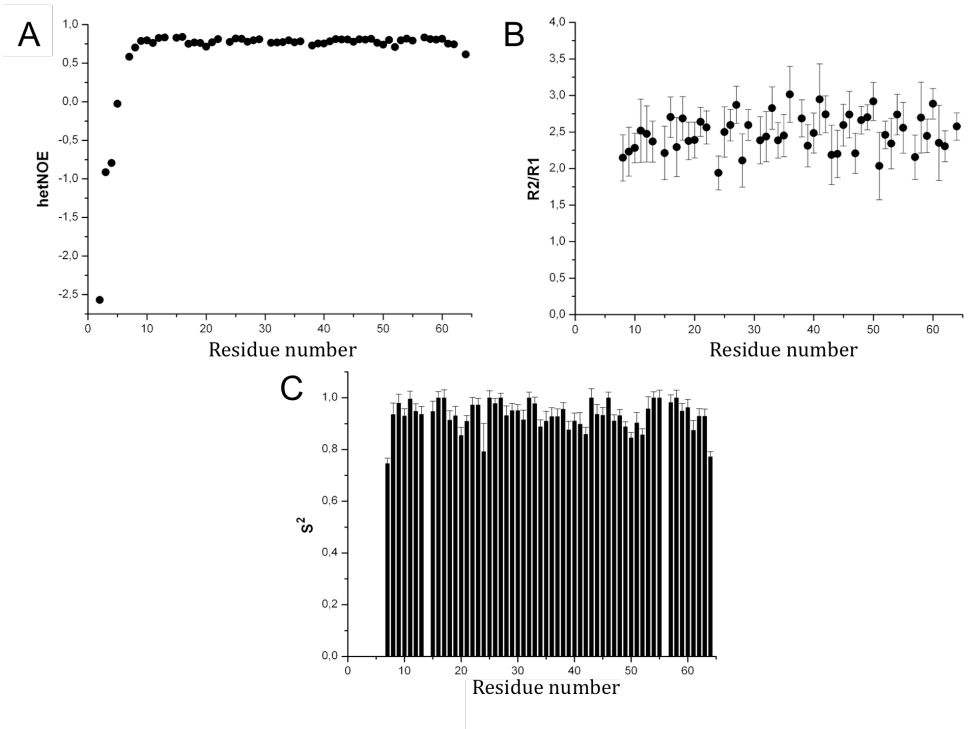


Figure 3. Backbone dynamics of CD2AP SH3-C measured at 600.25 MHz and 25 ºC. A. *^1H-^{15}N steady-state NOes (hetNOe).* ***B.*** *^{15}N R_2 over R_1 ratios (R2/R1); analysis of relaxation data using TENSOR2 show average R2/R1 ratios from residues in secondary structure elements that correspond to an apparent rotational correlation time t_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 t_m of 3.79 ns.*

DISCUSSION AND CONCLUSIONS

The three-dimensional structure of CD2AP SH3-C shows the 5 β-strands and the 3^{10} helix that are common in the family of SH3 domains (Figure 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 sequences that consists of a hydrophobic surface with three shallow pockets defined by the side chains of preserved aromatic residues that are indicated in Figure 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 7 N-terminal and 2 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 Figure 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.

ACKNOWLEGDEMENTS

This research is funded by grant BIO2005-04650 from the Spanish Ministry of Education and Science (MEC). 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.LR.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 and Wishart DS (2006) PREDITOR: a web server for predicting protein torsion angle restraints. Nucleic Acids Res 34: W63-69

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 and Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Cryst D54: 905–921

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J and 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 and 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 and 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 and 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 and Von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182: 319-326

Güntert P, Mumenthaler C and 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 and 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 and 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 and Bravo J (2005) Cbl promotes clustering of endocytic adaptor proteins. Nat Struct Mol Biol 12: 972-979

Koradi R, Billeter M and Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14, 51-5, 29-32.

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R and 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 Science 114: 1253-1263

Moncalián G, Cárdenes N, Deribe YL, Spínola-Amilibia M, Dikic I and 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 and 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 and 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 and 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

Chapter 3

Solution Structure, Dynamics and Thermodynamics of the three SH3 domains of CD2AP

Submitted to Protein Science (2010) by Jose L. Ortega-Roldan, Martin Blackledge, Nico A.J. van Nuland and Ana I. Azuaga

ABSTRACT

CD2 associated protein (CD2AP) is an adaptor protein that plays an important role in cell to cell union needed for the kidney function. It contains three N-terminal SH3 domains that are able to interact among others with CD2, ALIX, c-Cbl and Ubiquitin. To understand the role of the individual SH3 domains of this adaptor protein we have performed a complete structural, thermodynamic and dynamic characterization of the separate domains using NMR and DSC. The energetic contributions to the stability and the backbone dynamics have been related to the structural features of each domain using the structure-based FoldX algorithm. We have found that the N-terminal SH3 domain of both adaptor proteins CD2AP and CIN85 are the most stable SH3 domains in literature. This high stability is driven by a more extensive network of intra-molecular interactions. We believe that this increased stabilization of N-terminal SH3 domains in adaptor proteins is crucial to maintain the necessary conformation to establish the proper interactions critical for the recruitment of their natural targets.

INTRODUCTION

SH3 domains are probably the most common molecular recognition modules in the proteome [1, 2]. All SH3 domains share the same three-dimensional structure that consists of two orthogonal sheets of three anti-parallel beta-strands. The binding site for proline-rich ligands of these domains corresponds to a highly conserved hydrophobic region flanked by two loops (RT and n-Src) of more variable composition and mostly of polar nature, while ubiquitin interacts with a larger surface by tertiary contacts. A considerable amount of work has been carried out to characterize these domains in terms of their stability, folding thermodynamics and kinetics, three-dimensional structure, backbone and side chain dynamics and interactions with artificial ligands and natural targets [3-6]. Kinetic and equilibrium studies indicate that the folding-unfolding of these domains is a simple two-state process, without significantly populated intermediates. More recent results suggest that the SH3 domain shows an elevated, but not full structural cooperativity and therefore the conformational fluctuations of the different structural elements are partially coupled [7]. Such structural cooperativity implies that any local energetic change or interaction gets transmitted efficiently to the rest of the structure.

The majority of these studies have focussed on single SH3 containing proteins. In many proteins, however, SH3 domains occur in tandems of two or three units separated by short or long stretches of linking residues. Some of these proteins, such as those of the CIN85/CMS (Cbl-Interacting protein of 85 kDa/Cas ligand with Multiple SH3 domains) adaptor protein family,

show high homology among their SH3 domains suggesting similar specificities in ligand binding [8]. It has been reported previously that different domains are able to transmit communication along a protein via conformational changes [9, 10] or modification of internal dynamics [11].

CD2 associated protein (CD2AP) is an adaptor protein that belongs to the CIN85/CMS adaptor proteins family. It was first found in T-cells where CD2AP has been shown to be essential for CD2 clustering and for inducing T-cell polarization at the contact area between T-cell and antigen-resenting cell [12]. CD2AP knockout mice exhibit neprhotic syndrome characterized by defects in podocyte foot processes, suggesting that CD2AP plays an essential role in cell-to-cell union needed for the kidney function [13, 14]. Interactions of CD2AP with CD2 [12, 15], c-Cbl [15], ALIX [16] and Ubiquitin[17] among others are known to be mediated by its SH3 domains.

CD2AP is a cytosolic protein of 641 amino acids. The amino terminus consists of three SH3 domains (SH3 A, B and C) and a proline rich region, which is followed by a globular domain and a coiled-coil structure. Between the second and the third SH3 domains a serine/threonine rich region is found, which serves as a putative phosphorylation site. The proline rich region may act as a protein interaction site. In addition, the C-terminal binding motif for both filamentous and globular actin suggests that CD2AP may affect actin organization in the cell. At the C-terminus a coiled-coiled like structure is present, which may cause auto-dimerization of CD2AP.

As a first step toward understanding the role of the three CD2AP SH3 domains, we have carried out a complete structural, dynamic and thermodynamic characterization of the isolated domains. The thermal stability of each individual domain has been studied by Differential Scanning Calorimetry (DSC), allowing a systematic comparison between the stabilities of the different SH3 domains. The three dimensional solution structure of each individual domain was determined under the same experimental conditions permitting a structure-based explanation for the differences in stability found between the different SH3 domains. Moreover, the contribution of each residue to the global folding was studied by NMR Hydrogen Deuterium (H/D) exchange, and its relationship with backbone dynamics was evaluated with the information extracted from ^{15}N relaxation experiments.

Our results on the CD2AP SH3 domains reveal significant differences in the global folding parameters and in backbone dynamics in comparison with data from a variety of SH3 domains found in bibliography and predicted data from the structure of the SH3 domains of the homologous adaptor protein CIN85. These differences have been explained in terms of individual energetic contributions arising from their sequence and three dimensional structures, and rationalized in term of the biological role inside

the adaptor proteins, and constitute the basis for the study of full length SH3 containing tandems.

RESULTS

The solution structures of the three SH3 domains of CD2AP

In order to rationalize the stabilities of each of the three SH3 domains of CD2AP in terms of their structure, we have determined the high-resolution solution structures of each individual domain under the same experimental conditions at pH 6.0. All backbone resonances of SH3-A and C were assigned from 3D HNCACB, CBCA(CO)NH, HNCO and HBHA(CO)NH experiments (BMRB accession numbers 16643 and 16641, respectively). All proton resonances of SH3-B, except for the labile side-chain hydrogen atoms, were assigned from homonuclear 2D TOCSY and 2D NOESY ^{1}H NMR spectra (BMRB accession number 16642).

SH3-A: The X-ray structures of free and CD2-bound SH3-A domain (PDB entries 2J6K and 2J6O, respectively) were refined using 56 NH residual dipolar couplings measured in a PEG/n-hexanol mixture with the program SCULPTOR [18]. Imposing these NH RDCs on the two crystal structures resolved at 2.3 and 2.0 Å resolution [15], respectively, lead to the same unique structure (PDB 2KRM) (Figure 1A); the backbone RMSD for residues 3-56 between the lowest-energy structure of each set is 0.32 Å, the same as the variation within each structural set (RMSD of 0.25 Å and 0.32 Å, respectively; see table 1), confirming the effectiveness of the protocol used. Figure 2 shows that the calculated RDCs from the respective X-ray structures differ from the experimental determined RDCs in a similar manner.

SH3-B: The solution structure of the SH3-B domain was obtained using unambiguously assigned NOes from the automatic NOe assignment protocol in CYANA [19, 20] in a subsequent water-refinement with RECOORD [21] (Deposited in the Protein Data Bank with accession number 2KRN). Even though the structure is obtained from only distance restraints derived from homonuclear NOe spectra, the ensemble shows very good Ramachandran statistics while fulfilling the experimental data (Table 1), reflecting the high quality of the 3D structures. The average backbone RMSD for residues 6-58 is 0.77 ± 0.10 Å, which is similar to the 0.75 ± 0.14 Å found for the same residues in the ensemble of 10 lowest-energy NMR structures of the highly homologous SH3-B from the human equivalent CMS (PDB entry 2FEI; see Figure 1B), which was obtained from heteronuclear NMR data. Moreover, these two SH3-B structures are similar with a backbone RMSD of 1.09 Å for the same residues between the lowest-energy structures of the two NMR ensembles.

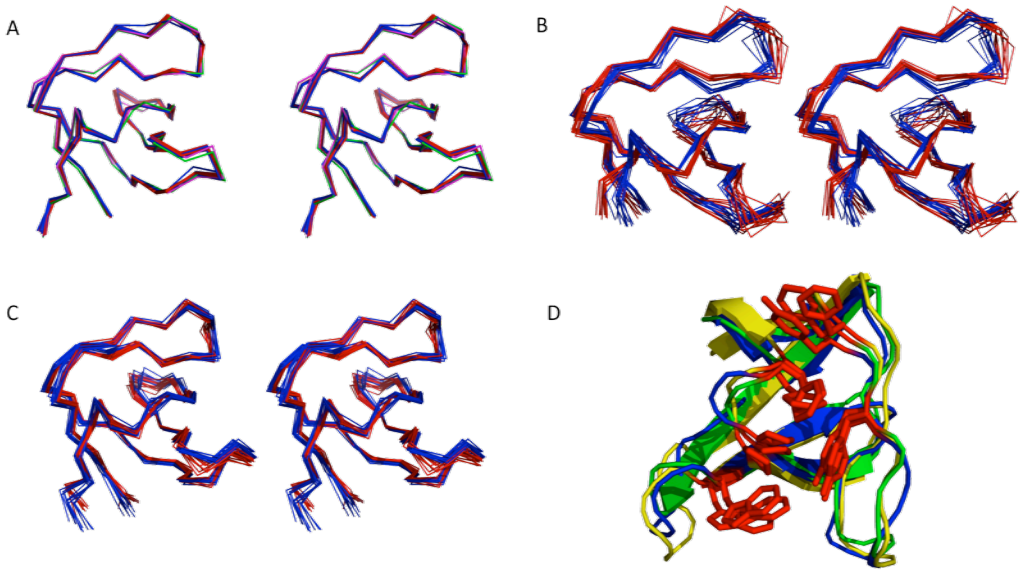


Figure 1: *(A) Stereo ribbon representation of the ensemble of the ten lowest energy solution structures of CD2AP SH3-A. Structures resulting from 2J6K refinement are coloured blue, structures from 2J6O refinement red. For comparison, the X-ray structures of CMS SH3-A in the free (PDB entry 2J6K) and CD2-bound form (PDB entry 2J6O) are shown in green and magenta, respectively. Structures were superimposed on the backbone atoms of residues 3-58. (B) Stereo ribbon representation of the ensemble of the ten lowest energy structures of CD2AP SH3-B (blue) and CMS SH3-B (PDB entry 2FEI, red) superimposed on the backbone atoms of residues 6-60 (3-57 in CMS SH3-B) (C) Stereo ribbon representation of the ensemble of the ten lowest energy structures of CD2AP SH3-C at pH 2 (red) and the RDC-refined at pH 6 (blue) superimposed on the backbone atoms of residues 6-64 (D) Cartoon representation of the lowest energy structure of CD2AP SH3-A (blue), SH3-B (green) and SH3-C (yellow). The residues involved in polyproline binding are presented in red sticks.*

SH3-C: The SH3-C structure was obtained from the refinement of the SH3-C pH 2.0 NMR structure (PDB entry 2JTE) using $^{1}D_{NH}$, $^{1}D_{C_\alpha C'}$, $^{2}D_{HN\ C'}$ and $^{1}D_{C_\alpha H_\alpha}$ RDCs measured in a PEG/n-hexanol mixture with the program SCULPTOR [18] (Deposited in the Protein Data Bank with accession number 2KRO). The ensemble fulfils the experimental data (Table 1, Figure 1C), with excellent Ramachandran statistics, reflecting the high quality of the 3D structures. The average backbone RMSD for residues 8-62 is 0.61 ± 0.10 Å, which is slightly higher than the 0.49 ± 0.09 Å found for the same residues in the ensemble of 10 lowest-energy NMR structures determined at pH 2 from only NOes (PDB entry 2JTE; see Figure 1C). The

higher RMSD found for the RDC refined structure at pH 6.0 is caused by the increased backbone dynamics present in the RT and n-Src loop and the C-terminus at this pH compared to pH 2.0 [22]. The structural differences between the two sets are small with a backbone RMSD of 0.94 Å for the same residues between the lowest-energy structures of the two NMR ensembles.

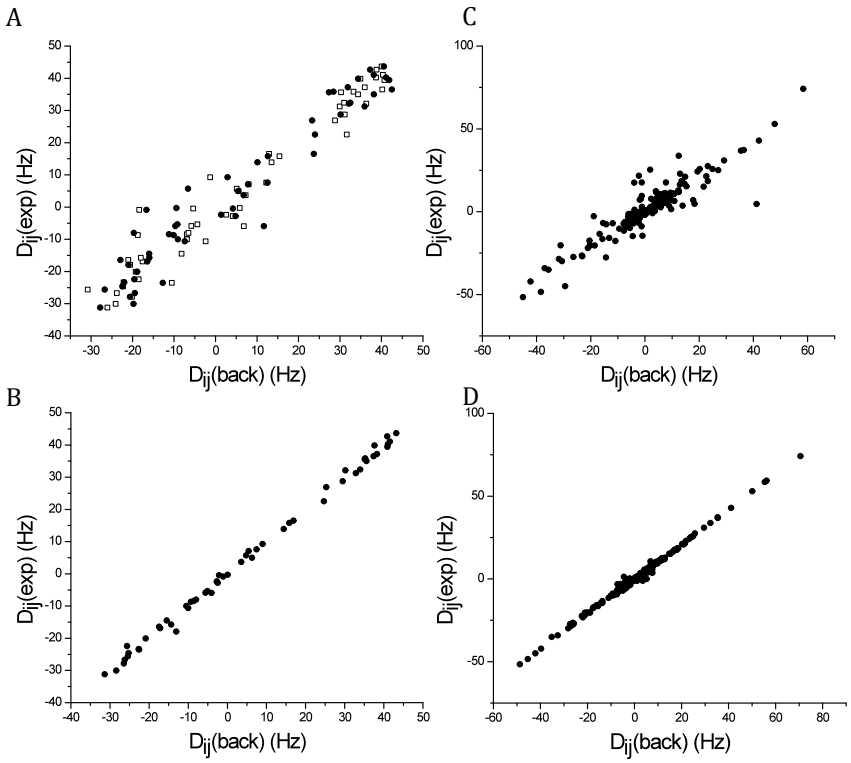


Figure 2: *Experimental versus back-calculated RDCs plots. (A) Experimental* $^{1}D_{NH}$ *RDCs measured vs back-calculated RDCs from PDB 2J6O (circles) and 2J6K (open squares). (B) Experimental* $^{1}D_{NH}$ *RDCs measured vs back-calculated RDCs from the lowest energy RDC-refined SH3-A structure (C) Experimental* $^{1}D_{NH}$, $^{1}D_{C_\alpha C'}$, $^{2}D_{HN\,C'}$ *and* $^{1}D_{C_\alpha H_\alpha}$ *RDCs measured for SH3-C vs back-calculated RDCs from the NMR structure of SH3-C at pH 2 (PDB 2JTE). (D) Experimental* $^{1}D_{NH}$, $^{1}D_{C_\alpha C'}$, $^{2}D_{HN\,C'}$ *and* $^{1}D_{C_\alpha H_\alpha}$ *RDCs measured for SH3-C vs back-calculated RDCs from the lowest energy RDC-refined SH3-C structure.*

Table 1: *Structural statistics for the 10 lowest energy structures of CD2AP SH3 A, B and C domains. Flexible N- and C-terminal residues were omitted from the RMSD analysis. NOe violations and Ramachandran statistics obtained from PROCHECK-NMR analysis over all residues.*

	SH3-A	**SH3-B**	**SH3-C**
Distance restraints			
Short range (i-j = 0)	-	225	237
Medium Range (1≤|i-j|≤4)	-	304	403
Long range (|i-j|≥5)	-	312	394
Orientation restraints			
Residual dipolar couplings	56	0	227
Restraint statistics			
NOe violations > 0.5 Å	-	0	0
RDC violations > 3 Hz	1	-	1
RDC violations > 5 Hz	0	-	0
RMSD from average (Å)	Residues 3-56	Residues 6-58	Residues 8-62
Backbone N, CA, C´	0.28 ± 0.06	0.77 ± 0.10	0.61 ± 0.10
Heavy atoms	1.62 ± 0.14	1.48 ± 0.10	1.29 ± 0.16
Ramachandran plot			
Most favored regions (%)	86.5	84.9	86.5
Additional allowed regions (%)	13.5	13.1	11.5
Generously allowed regions (%)	0.0	1.2	0.0
Disallowed regions (%)	0.0	0.8	1.9

Comparison between the three SH3 domains of CD2AP

The three SH3 domains of CD2AP studied here are highly homologous amongst themselves (Figure 3). Sequence alignment shows highest similarity between the SH3-A and C domains, with 53.2% of identity, followed by a 51% identity between SH3-B and C and a 46.2% identity between SH3-A and B. They share the typical aromatic residues involved in binding to proline rich sequences, but differ in the length of the n-Src loop where SH3-C has one or two additional residues compared to the A and B domain, respectively. The CD2AP SH3 domains do not only show high sequence identity among themselves, but sequence alignment with other SH3 domains (Figure 3) reveals high homology between the CD2AP SH3

domains and adaptor protein CIN85, suggesting targeting of similar natural partners [8].

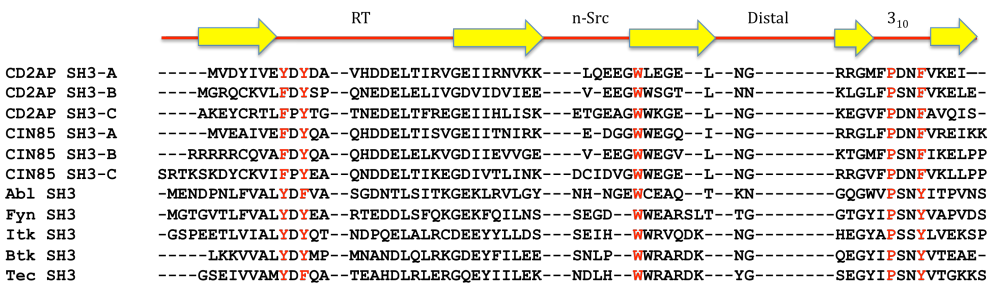


Figure 3: *Sequence alignment of the CD2AP SH3 domains with other related and non-related SH3 domains. Residues involved in binding to proline-rich sequences are shown in red.*

The three CD2AP SH3 domains show the five β-strands and the 3^{10} helix that are common in the family of SH3 domains (Figure 1D). They also contain all the features that are typically found in other SH3 domains, including the canonical binding site residues for the recognition of polyproline sequences that consist of a hydrophobic surface with three shallow pockets defined by the side chains of preserved aromatic residues that are indicated in figures 3 and 1D. The high sequence similarity is reflected substantial structural homology, with backbone RMSD of 1.36 Å and 1.20 Å for residues 3-32 and 37-56 in SH3-A and residues 6-35 and 39-58 in SH3-B compared to the equivalent residues 8-37 and 43-62 in SH3-C, and 1,29 Å between SH3-A and SH3-B for the same residues. The flexible N and C-termini, and the n-Src loop due to its different length in the three CD2AP SH3 domains were excluded in this analysis.

Figure 4 shows the surface-charge distribution of the three SH3 domains. A large polar surface area is found in each domain. SH3-A and SH3-B share a very similar electrostatic distribution with positively and negatively charges distributed similarly over the whole structural surface. In the polyproline binding area a mainly non-polar surface is found in SH3-A and SH3-B, flanked by positive and negatively charged residues. In contrast SH3-C displays a different electrostatic distribution in this area, presenting a highly negatively charged surface. The rest of the domain shows a similar charge distribution as in the other two domains.

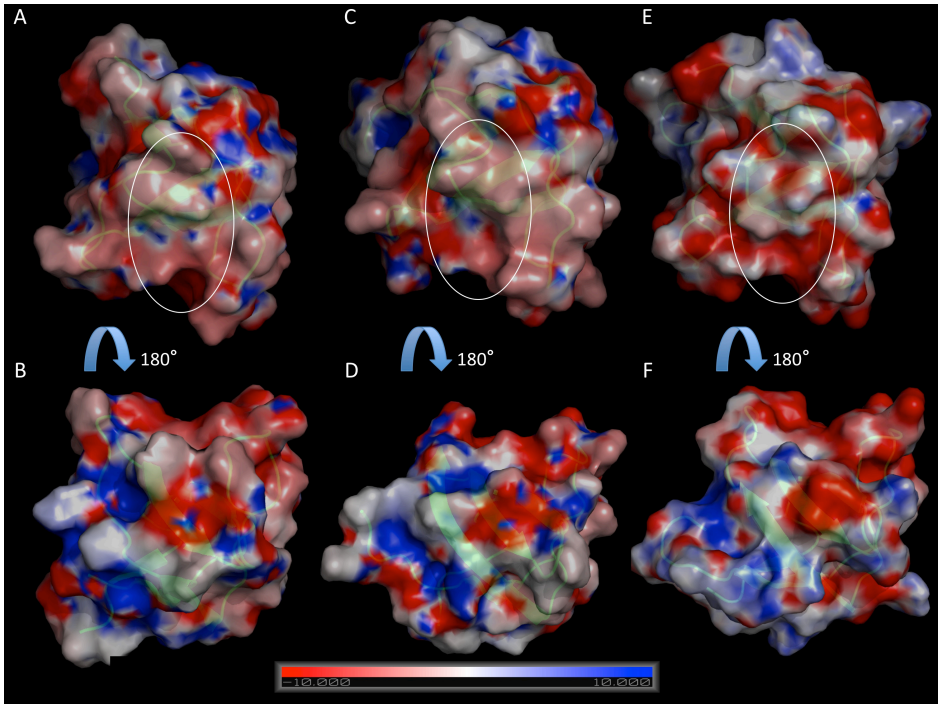


Figure 4: *Surface representation in 180° rotated orientations of CD2AP SH3-A (A and B), SH3-B (C and D) and SH3-C (E and F) domains, showing the surface-charge distribution. The electrostatic surface representation of the three structures was drawn using Pymol(www.pymol.org) using a Poisson-Boltzman electrostatics calculation[23]. Orientations in A, C and E are similar than in figure 1D. The polyproline binding pocket is indicated by a circle in each domain.*

In order to relate the structural features of the three domains with their thermodynamic behaviour, the optimal hydrogen bonding network and salt bridges were calculated for each structure using the Whatif WEB interface [24]. Table 2 shows the couples of amino-acids involved in salt bridges. Especially SH3-A exploits a high number of salt bridges, which is the result of the equal distribution of positive and negative charged residues over the whole structure as visualized in figure 4. At the same time a different number of hydrogen bonds are found in the three domains, varying from namely 35 found in SH3-A, 39 in SH3-B and 33 in SH3-C. Particularly remarkable is the stabilization of the C-terminus of SH3-C due to a hydrogen bonding and salt bridge network via the β-sheet formed by the β_1, β_2 and β_5 strands. Such network is absent in SH3-A, while direct interactions between the C- and N- terminus are present in SH3-B.

***Table 2**: Couples of amino-acids involved in salt bridges in CD2AP SH3-A, B and C.*

SH3-A	SH3-B	SH3-C
D3 – R27	E18 – K47	E7 – K36
E7 – K 55	E22 – K47	E27 – R10
D9 – R21	E58 – R5	E47 – H32
D16 – H14		E52 – K45
D16 – R45		
E17 – H14		
E29 – R21		
E39 – R46		
E41 – R27		
E41 – R46		
D51 – K31		
E56 – K58		

Thermal stability studied by DSC.

DSC is the most appropriate technique for the thermodynamic characterization of thermal transitions in proteins. The global stability of the three SH3 domains of CD2AP was measured by DSC at different pH values in order to explore, in combination with the NMR-detected H/D exchange, the nature of the native state conformational ensemble of the three domains. Prior to the DSC experiments the association state of every sample was analyzed with Dynamic Light Scattering (DLS) at 25°C using a concentration of 2 mg/mL to discard the presence of associated states. A hydrodynamic radius of 1.6 nm was obtained for all CD2AP SH3 domains, a value typical for a globular monomeric protein of this size [25, 26]. For the SH3-B domain, however, a small fraction of high-order associated states was found.

SH3-A: DSC experiments of the CD2AP SH3-A domain were carried out at pH values 2.0, 3.0, 6.0, and 7.0. For the pH range between 2.0 and 3.0 and between 6.0 and 7.0 the thermal unfolding thermal transitions were highly

reversible and independent of scan rate and of concentration. DSC experiments could not be performed at pH 4.0 due to aggregation of the sample even during dialysis, likely caused by approaching the theoretical isoelectric point (pI), estimated at 4.42 with the ExPASy ProtParam tool. In all other conditions where the transition was found to be reversible, molar heat capacity curves were fitted by a nonlinear least-squares method using the two-state unfolding model as described previously [6]. Table 3 summarizes the thermodynamic data at all pH values studied. Both the enthalpy and Gibbs energy of unfolding reach a maximum at pH 6.

SH3-B: DSC experiments were performed at pH 5.5, 6.5 and 7.4. Due to massive aggregation in the dialysis tube, no calorimetric study was possible in the pH range 2.0 to 4.0. Only at pH 5.5 the unfolding thermal transitions were partially reversible, where the reheating experiment shows two transitions, one centred at 53.7°C corresponding to the unfolding of the monomeric domain, and a high temperature second transition centred at 85°C. The latter transition can be explained as the unfolding of high molecular species observed at low temperatures by DLS (data not shown). No concentration or scan rate effect is present on the first scan of the sample, and the molar heat capacity curves accounted very well by the two-state unfolding model. At pH 6.0 and 7.4 the unfolding thermal transitions were completely irreversible. In this case, no reversible thermodynamic model can be applied and the calorimetric ΔH was estimated by integrating the area under the unfolding thermal transition and T_m was determined from its maximum.

SH3-C: DSC experiments were carried out at pH 2.0, 2.5, 3.0, 4.0, 6.0 and 7.0. In the pH range between 2.0 and 4.0 all unfolding thermal transitions were mostly reversible and independent of concentration and scan rate, although the reversibility is pH dependent, and decreases upon increasing the pH to values close to the theoretical isoelectric point (pI 4.6). However the reversibility in this pH range is always higher than 80%. For these conditions the molar heat capacity curves were fitted using a two-state unfolding model (Figure 5A). All calorimetric data are summarized in Table 3. An increase in stability is found when raising the pH from 2.0 to 4.0. At pH values higher than 4 the thermal unfolding is irreversible and only T_m and the calorimetric enthalpy can be calculated as for SH3-B. DLS experiments were performed at different temperatures to check the presence of associated states at high and low temperatures. At pH 2.0 no aggregates are present at any temperature (Figure 5C), consistent with the high reversibility of the thermal unfolding transition measured by DSC. The shift of the peak centred at 1.6 nm when raising the temperature to 2.1 nm is typical for denaturation of globular monomers into an unfolded polypeptide chain of this size [25, 27]. At pH 4.0 the thermal unfolding followed by DLS shows the appearance of states with higher hydrodynamic radius when

raising the temperature above 45°C, consistent with the decrease in reversibility found at this pH observed by DSC (Figure 5D).

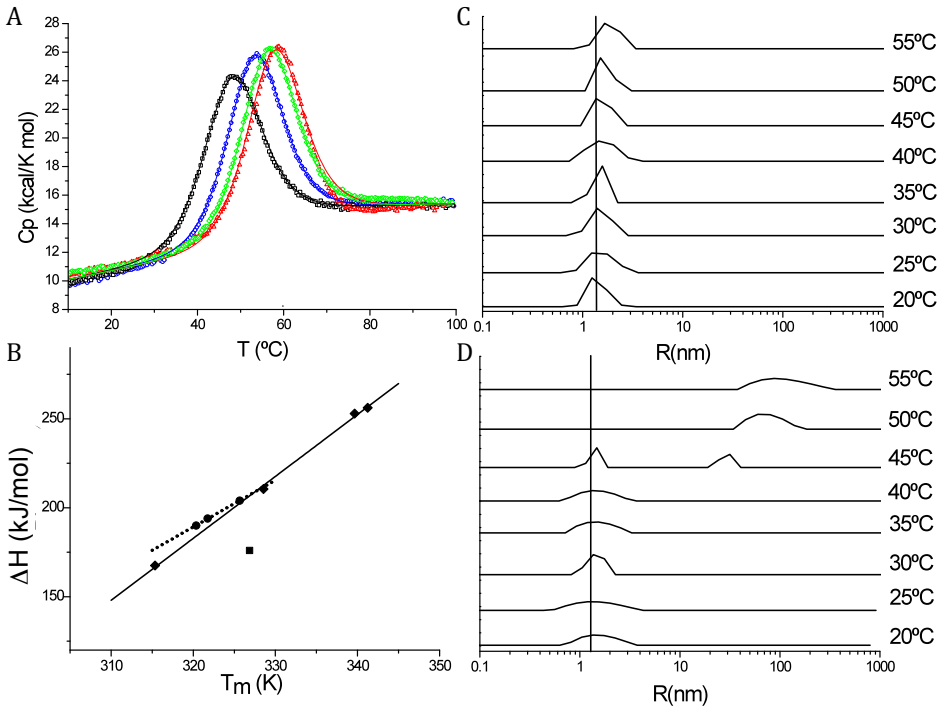


Figure 5. *(A) DSC thermograms of CD2AP SH3-C at pH 2.0 (black), 3.0 (blue), 3.5 (green) and 4.0(red). Solid line shows best fits using a two-state unfolding model. (B) ΔH vs T_m plot and linear fittings for SH3-A (◆ and black line), SH3-C (● and light dotted line) and ΔH vs T_m at pH 5.5 for SH3-B (■). (C) Size distribution obtained by DLS at different temperatures for SH3-C at pH 2.0. Vertical line centred at a hydrodynamic radius of 1.6 nm. (D) Size distribution obtained by DLS at different temperatures for SH3-C at pH 4.0. Vertical line centred at a hydrodynamic radius of 1.6 nm.*

Comparison of the thermodynamic properties of the three domains indicates a different behaviour of SH3-B compared with SH3-A and C caused by its high tendency to aggregate. At pH values below the pI SH3-A and C show a similar increase in stability and decrease in reversibility upon increasing the pH. At pH values above the pI, a small decrease in stability is found in SH3-C coupled with a high irreversibility in the thermal unfolding, while SH3-A shows a reversible thermal unfolding and much higher stability reflected in ΔG values 9-10 kJ/mol higher in the range of pH from 6.0 to 7.0.

Table 3. *Thermodynamic data obtained from DSC experiments for CD2AP SH3 domains at different pH values. (I) ΔH estimated by integrating the area under the unfolding thermal transition. (II) ΔC_p of unfolding calculated from the slope of ΔH vs T_m (Figure 5B).*

	pH	T_m (°C)	ΔH (kJ/mol)	ΔG (25 °C) (kJ/mol)	$\Delta C_p^{(II)}$ (kJ/mol·K)
SH3-A	2.0	42.2 ± 0.8	168 ± 3	7.7 ± 0.5	3.5 ± 0.1
	3.0	55.4 ± 0.8	211 ± 3	15.1 ± 0.6	
	6.0	68.1 ± 0.8	256 ± 3	22.4 ± 0.8	
	7.0	66.5 ± 0.8	253 ± 3	23.0 ± 0.7	
SH3-B	5.5	53.7 ± 0.9	176 ± 4	12 ± 2.2	2.3 ± 0.2
	6.5	53.2 ± 1	$154^{(I)}$ ± 10	-	-
	7.0	53.4 ± 1	$194^{(I)}$ ± 10	-	-
	7.4	51.8 ± 1	$207^{(I)}$ ± 10	-	-
SH3-C	2.0	47.2 ± 0.9	190 ± 4	11.2 ± 0.7	2.47 ± 0.1
	2.5	48.6 ± 0.9	194 ± 4	12.0 ± 0.7	
	3.0	52.5 ± 0.9	204 ± 4	14.2 ± 0.8	
	3.5	57.0 ± 0.9	208 ± 4	16.2 ± 0.8	
	4.0	58.7 ± 0.9	213 ± 4	17.2 ± 0.9	
	6.0	54.6 ± 0.9	$220^{(I)}$ ± 10	16.4 ± 1.3	
	7.0	50.4 ± 0.9	$173^{(I)}$ ± 10	11.0 ± 1.1	

Conformational flexibility and backbone dynamics from Hydrogen Deuterium exchange and ^{15}N relaxation data.

NMR detected H/D exchange experiments represents a unique tool for the study of protein folding, as it provides a means of monitoring the local unfolding reactions of each residue. H/D experiments were performed in order to relate the global unfolding parameters obtained from DSC with the structural features of CD2AP SH3-A and C domains. No data could be extracted for SH3-B due to aggregation during the preparation of the sample.

H/D exchange rates have been measured at pH 6.0 for SH3-A and C. Reliable data could be extracted for 37 residues out of 62 in SH3-A and 38 out of 64 residues in SH3-C. In all cases the exchange shows first-order kinetics and allows an accurate determination of the exchange rate constants, k_{ex}, assuming an EX2 mechanism. The experimental ΔG_{ex} per residue are shown in Figure 6C. In both SH3 domains, protected residues are grouped mainly in six regions of the protein sequence, corresponding to different elements of secondary structure.

In the case of SH3-A the protected regions correspond to (a) residues 4-13, located in the first β-strand and the first arm of the RT loop; (b) residues 17-25, in the second arm of the RT loop and the second β-sheet; (c) residues 39-51 positioned in the second half of the third β-sheet and the fourth β-sheet, except for the glycine 44 located in the edge of the distal loop; and (d) residues 51-57, corresponding to the 3^{10} helix and the fifth β-sheet except the aspartic 52 (Figure 6C). For this domain the RT and distal loop show measurable protection, including histidine 14, located at the edge of the RT loop. In contrast, the n-Src loop does not show measurable protection. One non-observable proline residue is present in position 50. Comparison of the protection pattern between the different secondary structure elements shows high and similar ΔG_{ex} values except for the C-terminal region (residues 51-57), including the 3^{10} helix and the fifth β-sheet. The core of the protein is therefore constituted by the β_1, β_3 and β_4 strands and the two arms of the RT loop, as they present individual ΔG_{ex} of the same order than the ΔG obtained by DSC. On the other hand, lower ΔG_{ex} values are found in β_2 and β_5 strands and the 3^{10} helix indicating a higher flexibility of these areas (Figure 6A).

In the case of the SH3-C domain, the protected regions correspond to: (a) residues 8-15 in the first β-sheet and the first arm of the RT loop; (b) residues 22-34, located in the second arm of the RT loop and the second β-sheet; (c) residues 43-48, in the third β-sheet; (d) residues 51-55, positioned in the fourth β-sheet; and (e) residues 57-61, corresponding to the 3^{10} helix and the fifth β-sheet, excluding the asparagine 58 (Figure 6C). As in SH3-A the n-Src loop does not show measurable protection. In contrast, RT and distal loops show much lower protection to exchange in SH3-C. Two non-observable proline residues are at positions 14 and 56. For this domain, all secondary structure elements show comparable ΔG_{ex} values and similar to the ΔG obtained by DSC and thus form the cooperative core of the protein (Figure 6B).

^{15}N relaxation measurements give information on the ps to ns and μs to ms timescale of the backbone dynamics, complementing the ms to s timescale dynamics provided by the H/D exchange. The molecular rotational correlation time t_c was calculated for SH3-A and SH3-C using the program TENSOR2 [28] from selected R2/R1 values of residues that are not affected by extensive disorder on a ps time scale (low NOe) or chemical exchange on a μs-ms time scale (elevated R2/R1 ratio), corresponding to residues 4-10, 26, 28-30, 44-49, 55-56 in SH3-A and residues 9-12, 28-33, 45-47, 52, 56, 60-62 in SH3-C. Molecular rotational correlation times of 4.24 and 3.55 ns were obtained for SH3-A and C respectively, corresponding to a radius of hydration of 17.1 Å and 16.1 Å, values expected for globular monomeric molecules of this size [25-27]. Moreover, these values are in good agreement with the hydrodynamic radius obtained with DLS at much lower

concentrations, indicating no association phenomena in this range of concentrations. The relaxation data for both domains are shown in Figure 7.

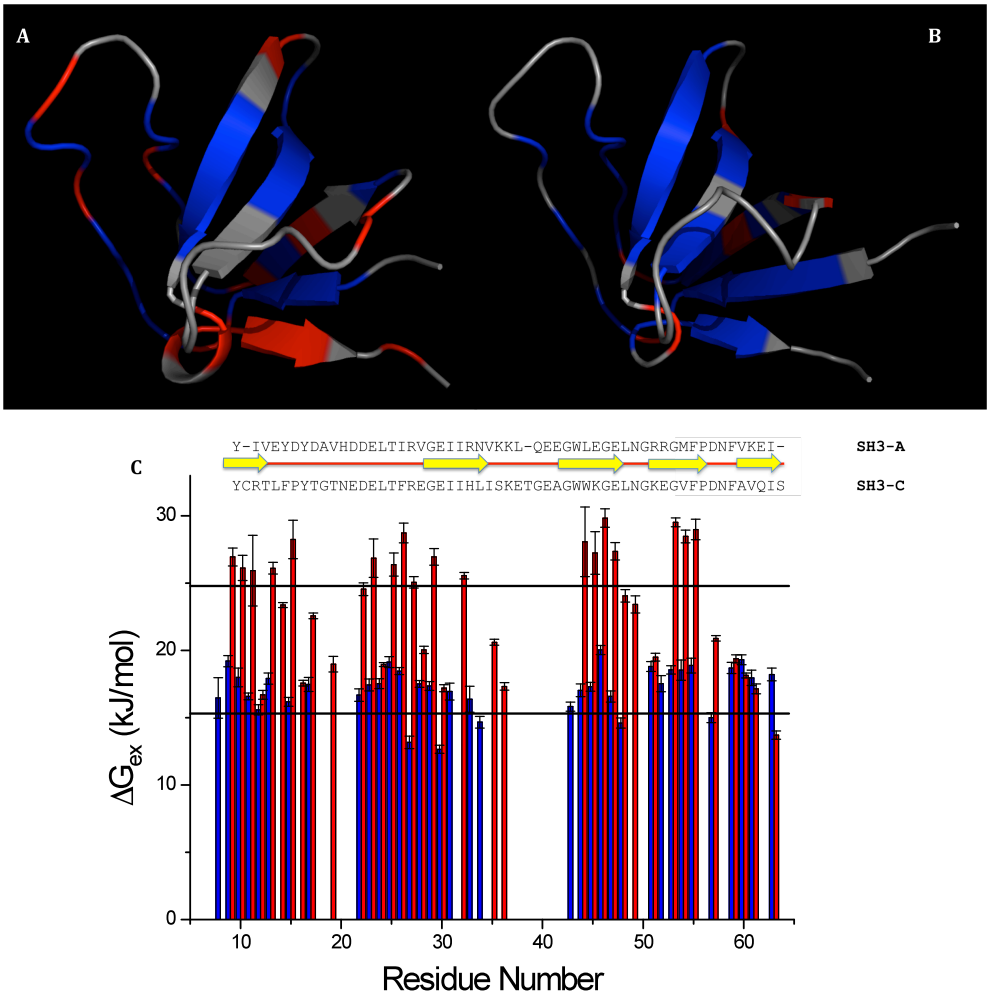


Figure 6. *Plot of the cooperative core and the flexible regions undergoing partial unfolding on the structures of CD2AP SH3-A (A) and SH3-C (B). Residues with ΔG_{ex} higher than the ΔG obtained by DSC minus three standard deviations (3std) are represented in blue. Residues with ΔG_{ex}<ΔG-3std are represented in red. Residues exchanging within the dead time of the experiment are represented in grey. (D) ΔG_{ex} calculated from H/D exchange for SH3-A (red) and SH3-C (blue), superimposed on the SH3-C sequence. Black lines indicate the ΔG obtained by DSC for SH3-A and C domains (Upper and lower line respectively).*

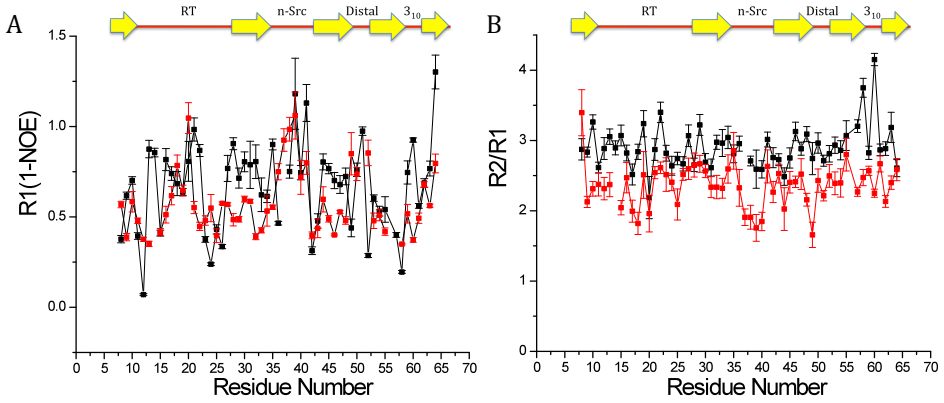


Figure 7. *Backbone dynamics of CD2AP SH3-A (black) and SH3-C (red) measured at 600.25 MHz and 25°C. (A) R1 (1-hetNOe), (B) ^{15}N R2/R1. For comparison, all relaxation values are shown according to the residue numbers in SH3-C. The secondary structure elements found in both domains are shown at the top of the figure.*

Generally, R1(1-hetNOe) values, which give information about extensive high-frequency motions on the ps-subns timescale, are similar for both domains, while the R2/R1 ratio are higher for the SH3-A domain. The average R2 values for SH3-A and SH3-C are 7.0±0.8 s^{-1} and 5.6±0.6 s^{-1} in the secondary structural elements, while average R1 values are 2.4±0.1 s^{-1} and 2.3±0.1 s^{-1}. Despite the rather high scatter, both R1 and R2 values are systematically higher for SH3A, indicating that the change in R2/R1 ratio is not solely due to an increase in overall rotational correlation time, but is also indicative of an increase in low frequency motions that result in an increase in R1. We prefer to present the data in this format, rather than using a standard Lipari-Szabo type model-free analysis, because of the difficulty of separating contributions from overall rotational diffusion and internal motion when significant nanosecond motions appear to be present. This kind of problem is particularly acute for small molecules, where the overall rotational correlation coefficient is less then 5ns. The overall higher R2/R1 values obtained for SH3-A results in a higher τ_c for this domain. However, relative differences in these values are found along the sequence. SH3-C shows a decrease in R2/R1 values in the RT, n-Src and distal loops caused by an increase in R1, indicating fast motions in these regions. Interestingly this effect is not seen in R2/R1 values measured at pH 2.0 [22]. On the other hand high R2/R1 values are found in the C-terminal region of SH3-A, an indicative of chemical exchange on the μs to ms timescale.

DISCUSSION

SH3 domains are small modular domains involved in protein-protein interactions. The sequence identity between different SH3 domains is low, ranging from 15 to 45% [29], although all SH3 domains adopt a similar five-stranded β-barrel structure. In the present work, we have determined the three dimensional solution structures of the three SH3 domains of CD2AP at pH 6.0. A high sequence homology is found between them and also with other related adaptor proteins, such as CIN85, what is translated in a high structural homology (Figures 3 and 1D). Moreover, the aromatic residues that are involved in recognition of polyproline sequences (Figure 1D) are similarly positioned in the three CD2AP SH3 domains. A similar electrostatic distribution is found on the surface around the polyproline binding pocket in SH3-A and B, dominated by a large negative area in the specificity pocket and a non-polar surface in the hydrophobic pockets surrounded by positive and negatively charged residues, what is consistent with the overlapping specificities suggested elsewhere [8]. However, SH3-C presents a higher negatively charged surface around the hydrophobic pockets. Interestingly, differences are found between CIN85 SH3-C and CD2AP SH3-C in this area. CIN85 SH3-C displays a more positively charged region around the hydrophobic pockets, similar to the A and B domains of both CIN85 and CD2AP, indicating possible differences in binding affinities and specificity to natural targets.

Despite of the similarity between the three dimensional structures of the CD2AP SH3 domains, except for the n-Src loop orientation, differences are found in the number and distribution of intra-molecular interactions. A computational structure based energy calculation using the program FoldX [30] has been carried out in order to relate the contribution of the interactions to the stability with the experimental thermodynamic parameters obtained by DSC. The contribution of these intra-molecular interactions to the total ΔG is reflected in the enthalpy change upon unfolding (ΔH). This value is dominated by two terms: a positive contribution due to the disruption of internal interactions (hydrogen bonds, van der Waals, salt bridges, etc.) and a negative term resulting from the solvation of groups that are buried in the native state [31]. FoldX predicts a higher enthalpy value for SH3-A domain driven by a 1 kJ/residue higher hydrogen bond energy than SH3-C and a 0.7 to 1.57 kJ/res higher van der Waals energy than SH3-B and C. Nevertheless, similar hydrogen bond energy contribution to the ΔH is found in SH3-B. This result is consistent with the higher number of hydrogen bonds predicted by Whatif for SH3-A and B. On the other hand, the experimental enthalpy values obtained for SH3-B are lower than the predicted ones, probably due to inaccuracy in the experimental value caused by the presence of high molecular weight species even at room temperature at neutral pH values, as it has been observed in DLS experiments. As no major differences are found in the entropy term

calculated from thermodynamic data for the three CD2AP SH3 domains, the 0.2 kJ/res higher stability of SH3-A at 25°C is driven by the enthalpic term.

The global folding data obtained by DSC has been compared with local folding information obtained by NMR detected amide Hydrogen/Deuterium exchange in SH3-A and C domains. No information could be extracted for SH3-B due to its high tendency to aggregate. The stable cores of both SH3 domains, defined as those regions where individual ΔG_{ex} are of the same order than the ΔG obtained by DSC, comprise most of the secondary structural elements, except for the C-terminal part of SH3-A where lower ΔG_{ex} values are found. The amide exchange of these core residues thus occurs primarily via a global unfolding reaction, while in the β_5 strand and the C-terminus in SH3-A, however, H/D exchange takes place via local or subglobal unfolding processes, indicating higher flexibility for this region of the protein. This effect is also observed at a different timescale as can be appreciated from the higher ^{15}N relaxation R2/R1 values of residues in this region, indicating that this area is affected by low frequency motions. This is not the case for SH3-C, where similar R2/R1 values are found for this part of the protein as in the rest of the domain in agreement with the presence of hydrogen bonds stabilizing the C-terminus that are absent in domain A. The increased mobility in the C-terminus of the SH3-A domain allows a higher flexibility in the junction with the rest of CD2AP that may permit regulation of interactions with natural targets to the SH3-B or C domains or to other regions of the protein.

The higher number of hydrogen bonds and salt-bridges present in SH3-A compared to SH3-C does not only result in a higher stability at 25°C but also affects the backbone flexibility by the different distribution of the intramolecular interactions along the sequence. The differences in local motions are mainly reflected by the differences in ^{15}N relaxation R2/R1 values [32]. Lower amplitude motions in the high frequency timescale are found for the RT, n-Src and Distal loops in SH3-A, consistent with a higher number of hydrogen bonds and salt-bridges calculated in these regions from the structure. This extended hydrogen bonding network is confirmed by H/D exchange data, where measurable protection is found for His 14 in the RT loop, Val 29 and Lys 30 in the n-Src loop, and Gly 44 in the Distal loop of SH3-A. The lower flexibility exhibited in these loops in SH3-A is in contrast to that observed in SH3-C and other SH3 domains, where increased flexibility is generally found for RT, n-Src and Distal loops [33-35]. Thus the lower flexibility of the backbone of CD2AP SH3-A domain is related with the higher overall stability resulting from the more extended intramolecular interaction network along the structure. Furthermore, comparison of R2/R1 values for SH3-C measured at pH 2.0 and 6.0 reveals similar motions for SH3-C at low pH than SH3-A at pH 6.0, in agreement with the similar thermal stability observed at low pH values for SH3-A and

C. These results suggest that backbone flexibility of the SH3 domains closely correlates with the overall thermal stability as observed for other systems [32].

The stability of the CD2AP SH3 domains has been compared with the thermal stability of other non-related SH3 domains of Abl, Fyn, Itk, Btk and Tec [3, 36]. While CD2AP SH3-B and C show experimental ΔG from DSC in the same order than the compared SH3 domains, SH3-A shows an average 0.20 kJ/res higher experimental stability. No major differences in charge distributions are observed between CD2AP SH3 domains and the Fyn, Abl, Itk, Btk and Tec SH3 domains, indicating that differences in thermal stability do not arise from differences in the electrostatic distribution in the surface of the domains. CD2AP SH3-A displays up to 1.5 kJ/res higher experimental ΔH than Fyn, Abl, Itk, Btk and Tec SH3 domains. In order to explain these entalphy differences, each individual contribution to ΔH have been calculated from the structure using the program FoldX. For all non-related SH3 domains an average 1.07 kJ/res lower hydrogen bond energy term is found, what justify the experimental ΔH differences. Differences in the experimental values of ΔC_p are also found. CD2AP SH3 domains display up to 0.8 kJ/K mol lower values than the mean ΔC_p found in literature for SH3 domains. ΔC_p differences arise from the different changes in average polar and non-polar surface accessibilities between the folded and the extended unfolded state. The difference in total, non-polar and polar surface accessibility is bigger in the Fyn, Abl, Itk, Tec and Btk SH3 domains than in the CD2AP SH3 domains, resulting in lower predicted ΔC_p values for the CD2AP SH3 domains, thus in agreement with the experimental results[37]. The higher thermal stability showed by the CD2AP SH3-A compared to non-related SH3 domains can be explained then by a higher ΔH caused by a more extensive intramolecular interaction network as observed by H/D exchange and ^{15}N relaxation measurements.

No experimental thermodynamic data can been found in literature for adaptor proteins containing multiple SH3 domains in tandem. In order to compare the thermodynamic data obtained for CD2AP with other adaptor proteins with similar domain structure, the SH3 domains of the related CIN85 were analyzed using FoldX. We used the structures of the three CIN85 SH3 domains determined by either NMR or X-ray (PDB entries 2BZ8, 2O6O and 2K9G for SH3-A, B and C, respectively). CIN85 shows a very high sequence homology with CD2AP, ranging from 68.4% of homology for both N-terminal SH3 domains to 67.3 % of homology between central SH3 domains and 58.2% of homology between the C-terminal SH3 domains (see Figure 3). In both proteins, the N-terminal SH3 domain displays very similar hydrogen bonding, van der Waals and solvation energies, suggesting closely similar enthalpy values. Moreover, similar enthalpy values are predicted for the second and third SH3 domains

despite differences in energy contributions, with higher hydrogen bond energies but lower solvation energies. This analysis suggests that such an elevated stability for an N-terminal SH3 domain might be a general characteristic of multiple domain containing proteins. As stated before, the CD2AP SH3-A domain displays an average 0.20 kJ/res higher experimental ΔG than Abl, Fyn, Brk, Itk and Tec SH3 domains at physiological pH values. All these domains are located in the middle of the sequence of their full-length proteins, and in some cases, like Abl SH3 domain, the presence of a N-terminal domain is required to maintain its three dimensional structure [38]. The high ΔG observed in CD2AP SH3-A and predicted for CIN85 SH3-A suggest that N-terminal SH3 domains need a very high stability in order to maintain a correctly folded conformation to establish the proper high affinity interactions with their natural targets crucial for the recruitment process. Indeed, binding studies performed in our group confirm that all three SH3 domains bind to CD2, but with a significant higher affinity to the N-terminal domain (unpublished data).

We have shown the relationship between sequence, three dimensional structure, stability, and dynamics of single SH3 domains and rationalized it in terms of the function inside the adaptor proteins. The work presented here will form the basis for a better understanding on how interactions of each individual domain to their natural targets are affected by the presence of the others. Work on different CD2AP SH3 domain tandems is in progress in our group to address this question.

MATERIALS AND METHODS

Protein expression and purification

Unlabelled CD2AP SH3-A was obtained as described [15]. ^{15}N-labeled and ^{13}C, ^{15}N-labeled CD2AP SH3-A and SH3-C were obtained as described [22]. CD2AP SH3-B and SH3-C gene, originally cloned in the plasmid pGEX 4T1 with BamH-I and Xho-I, were sub-cloned in a pGAT-2 plasmid with Nco-I and Hind-III covalently linked to a N-terminal 6xHis tag, a GST tag and a Thrombin cleavage site. Unlabelled SH3-B and C were produced by growing on TB medium at 27 °C till an OD_{600} of 0.7. Protein expression was induced with 1 mM IPTG at 27 °C and cells were grown overnight. Cells were resuspended after centrifugation for 15 min at 10,845 g in 50 mM phosphate, 0.3 mM NaCl, pH 8.0 buffer (Column Buffer: CB) and broken in a French Press. After centrifugation at 104,630 g during 30 min, the supernatant was loaded on a 5 mL Ni-Sepharose resin (Pharmacia) equilibrated with CB. The protein was eluted with a gradient of 0 to 500 mM imidazole in CB. The GST and His tag were cleaved by incubation with Thrombin (Sigma). Thrombin was eliminated by filtration over a Glutamidine Sepharose resin (Amersham Bioscience). As a final purification step the cleaved SH3-B and C were loaded on a Superdex G75 gel-filtration column (Pharmacia). Sample concentration was determined spectrophotometrically using extinction coefficients of 9970, 12490 and 14980 M^{-1} cm^{-1} for SH3-A, B and C respectively, determined using the ProtParam tool (Expasy).

Differential Scanning Calorimetry

Temperature scans were performed at a protein concentration of 1 mg/mL in a VP-DSC microcalorimeter (Microcal, Northampton, MA, USA) between 5 and 100 °C at a scan rate of 1.5 °C/min. Unfolding experiments were carried out in 20 mM Glycine, 20 mM Sodium Acetate or 20 mM Sodium Phosphate buffers at different pH values from 2.0 to 7.4. The reversibility of the thermal unfolding was checked in a second consecutive scan of the same sample. Instrumental baselines were subtracted to the experimental thermograms of the samples and the time response of the calorimeter was then corrected. The partial molar heat capacity curves (C_p) were calculated and analyzed using Origin 7.0 (OriginLab, Northampton, MA, USA).

Dynamic Light Scattering

Dynamic light scattering (DLS) measurements were carried out on a DynaPro MS-X instrument (Wyatt Technology Corporation, Santa Barbara, CA, USA) using a 30 μL quartz thermostated sample cuvette at different temperatures. Samples were prepared in 20 mM Glycine, 20 mM Sodium Acetate or 20 mM Sodium Phosphate buffers at pH values from 2.0 to 7.0 using a protein concentration 2 mg/mL. The protein solutions and the buffer were centrifuged and filtered through 0.02 μm Anotop filters (Whatman plc, Brentford, Middlesex, UK) immediately before measurements. Sets of 20 accumulations of 20 s each were acquired. Laser power was adapted to avoid early saturation of the instrument. Dynamics software (Wyatt Technology Corporation, Santa Barbara, CA, USA) was used in data collection and processing.

NMR Spectroscopy

Assignment. ^{15}N-labeled and ^{13}C, ^{15}N-labeled samples of CD2AP SH3-A and C and unlabelled samples of SH3-B were prepared for NMR experiments at 1 mM in 92 % H_2O/8 % D_2O, 50 mM NaPi pH 6.0. 1mM DTT was added to SH3-B and C samples. NMR spectra were recorded on a Varian NMR Direct-Drive Systems 600MHz spectrometer (^{1}H frequency of 600.25 MHz). Assignment of SH3-A and C domains was carried out as previously described [22]. Due to the high tendency to aggregate of SH3-B (less than 2 days), we used the homonuclear assignment method [39] by recording two-dimensional NOESY and TOCSY spectra on an unlabelled sample of SH3-B with mixing times of 125 and 70 ms, respectively. Water was suppressed using the excitation sculpting method [40]. All NMR data were processed using NMRPipe [41] and analyzed by NMRView [42] or Sparky [43].

Measurement of Residual Dipolar Couplings. Residual dipolar couplings (RDCs) were measured in samples partially aligned in a liquid-crystalline medium consisting of a mixture of 5% penta-ethyleneglycol monododecyl ether ($C^{12}E^5$) and hexanol [44]. $^1D_{NH}$ RDCs were measured in the SH3-A domain from a pair of spin-state-selected ^{1}H-^{15}N correlation spectra recorded using an DIPSAP filter [45] for J-mismatch compensated spin-state selection, and the BEST concept for longitudinal relaxation and sensitivity enhancement [46, 47] on a Varian NMR Direct-Drive Systems 600 MHz spectrometer. $^1D_{NH}$, $^1D_{C_\alpha C}$, $^2D_{HNC'}$ and $^1D_{C_\alpha H_\alpha}$ RDCs were measured in the SH3-C domain as described [48].

RDC refinement of the SH3-A and C domains of CD2AP

The program SCULPTOR [18], which runs under CNS [49], was used for the refinement of the structures of SH3-A and C. The protocol involves a restrained MD calculation using the standard CNS force field. The solution structure of SH3-A was obtained by refining two different initial X-ray structures of the domain determined in the absence and presence of CD2 peptide (PDB 2J6K and 2J6O, respectively) [15]. The backbone conformation is initially restricted to its X-ray coordinates using a harmonic potential and a distance restraint file calculated from all H_{α}-HN distances lower than 10 Å in the SH3-A free X-ray structure and refined using $^{1}D_{NH}$ RDCs. In case of SH3-C the full set of RDCs was used to refine the initial pH 2.0 free SH3-C structure (PDB 2JTE). The backbone conformation was restricted in this calculation using the NOe-restraints obtained at pH 2.0. The five lowest-energy structures of each set of refined SH3-A domain and the ten lowest-energy structures of SH3-C domain were used for the final analysis.

Structure calculation of the SH3-B domain of CD2AP

All ^{1}H except for labile side-chain protons were assigned from 600 MHz 2D TOCSY and NOESY spectra. NOe cross peaks were obtained by automatic peak picking in the 2D NOESY followed by manual removal of diagonal peaks and peaks arising from artifacts (e.g. residual water). NOes were assigned using the automated NOe assignment procedure of CYANA version 2.1 [20, 50]. The standard protocol was used with seven cycles of combined automated NOe assignment and structure calculation of 100 conformers in each cycle. Unambiguously assigned restraints were used for a final structure refinement in explicit solvent using the RECOORD protocol, which runs under CNS [21]. The ten lowest-energy structures were used for final analysis.

Amide Hydrogen Deuterium exchange

Hydrogen/Deuterium (H/D) exchange experiments are often used to detect unfolding events at the residue level. Seventy microliters of concentrated samples (3,5 mM) of CD2AP ^{15}N labeled SH3-A and C were dissolved in 500 µL of $^{2}H_2O$ at pH 6.0. ^{15}N 2D SOFAST-HMQCs [51, 52] were acquired at 25 °C at different times until a decay of approximately 90 % of the slowest exchanging proton for SH3-A and C respectively. Spectra of 1024*32 complex points were recorded using a relaxation delay of 500 ms resulting in a total experimental time of 5’ 36’’ and 5’ 47’’ for SH3-A and C. Hydrogen exchange rates were determined by fitting the decay in cross-

peak intensities versus time to a single exponential using the program Origin 7.0. The Gibbs free energy ΔG_{ex} per residue was calculated from the intrinsic exchange rates [53] using the following equation:

$$\Delta G = -RT \ln (k_{ex}/k_{int}) = -RT \ln (1/PF)$$

Backbone dynamics from ^{15}N relaxation data

^{15}N 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-A and C. Relaxation values were obtained from series of 2D experiments with coherence selection achieved by pulse field gradients using the experiments described previously [54] on ^{15}N-labelled SH3-A and 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.

ACKNOWLEDGEMENTS

This research was funded by grant BIO2005-04650 from the Spanish Ministry of Education and Science (MEC) and FQM-02838 from the Andalucia Regional Goverment. We thank Manuel Iglesias-Bexiga and Maria-Luisa Romero-Romero for initial experimental work, Malene Ringkjobing-Jensen for help with the RDC measurement, Salvador Casares and Francisco Conejero-Lara for useful discussions and Obdulio Lopez-Mayorga for his continuous support. J.L.O.R. is supported by a FPU pre-doctoral grant from the MEC. The CD2AP SH3-A clone was kindly provided by Jeronimo Bravo. The 600 MHz spectra were recorded in the Centre for Scientific Instrumentation (CIC) of the University of Granada and at the RALF Large Scale Facility in Grenoble, which is funded by the 'Access to Research Infrastructures' program of the European Union. The CD2AP SH3-B structure calculation was carried out in the supercomputing facilities (UGRGrid) of the University of Granada.

REFERENCES

1. Mayer, B.J., *SH3 domains: complexity in moderation.* J Cell Sci, 2001. **114**(Pt 7): p. 1253-63.
2. Li, S.S., *Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction.* Biochem J, 2005. **390**(Pt 3): p. 641-53.
3. Filimonov, V.V., et al., *A thermodynamic analysis of a family of small globular proteins: SH3 domains.* Biophys Chem, 1999. **77**(2-3): p. 195-208.
4. Guijarro, J.I., et al., *Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy.* J Mol Biol, 1998. **276**(3): p. 657-67.
5. Plaxco, K.W., et al., *The folding kinetics and thermodynamics of the Fyn-SH3 domain.* Biochemistry, 1998. **37**(8): p. 2529-37.
6. Viguera, A.R., et al., *Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition.* Biochemistry, 1994. **33**(8): p. 2142-50.
7. Casares, S., et al., *Cooperative propagation of local stability changes from low-stability and high-stability regions in a SH3 domain.* Proteins, 2007. **67**(3): p. 531-47.
8. Dikic, I., *CIN85/CMS family of adaptor molecules.* FEBS Lett, 2002. **529**(1): p. 110-5.
9. Bendsen, S., et al., *The QTK loop is essential for the communication between the N-terminal atpase domain and the central cleavage--ligation region in human topoisomerase IIalpha.* Biochemistry, 2009. **48**(27): p. 6508-15.
10. Calvete, J.J., *Structures of integrin domains and concerted conformational changes in the bidirectional signaling mechanism of alphaIIbbeta3.* Exp Biol Med (Maywood), 2004. **229**(8): p. 732-44.
11. Lenaerts, T., et al., *Quantifying information transfer by protein domains: analysis of the Fyn SH2 domain structure.* BMC Struct Biol, 2008. **8**: p. 43.
12. Dustin, M.L., et al., *A novel adaptor protein orchestrates receptor patterning and cytoskeletal polarity in T-cell contacts.* Cell, 1998. **94**(5): p. 667-77.
13. Shih, N.Y., et al., *Congenital nephrotic syndrome in mice lacking CD2-associated protein.* Science, 1999. **286**(5438): p. 312-5.
14. Li, C., et al., *CD2AP is expressed with nephrin in developing podocytes and is found widely in mature kidney and elsewhere.* Am J Physiol Renal Physiol, 2000. **279**(4): p. F785-92.
15. Moncalian, G., et al., *Atypical polyproline recognition by the CMS N-terminal Src homology 3 domain.* J Biol Chem, 2006. **281**(50): p. 38845-53.

16. Usami, Y., S. Popov, and H.G. Gottlinger, *Potent rescue of human immunodeficiency virus type 1 late domain mutants by ALIX/AIP1 depends on its CHMP4 binding site.* J Virol, 2007. **81**(12): p. 6614-22.
17. Stamenova, S.D., et al., *Ubiquitin binds to and regulates a subset of SH3 domains.* Mol Cell, 2007. **25**(2): p. 273-84.
18. Hus, J.C., D. Marion, and M. Blackledge, *De novo determination of protein structure by NMR using orientational and long-range order restraints.* J Mol Biol, 2000. **298**(5): p. 927-36.
19. Guntert, P., *Calculating protein structures from NMR data.* Methods Mol Biol, 1997. **60**: p. 157-94.
20. Guntert, P., C. Mumenthaler, and K. Wuthrich, *Torsion angle dynamics for NMR structure calculation with the new program DYANA.* J Mol Biol, 1997. **273**(1): p. 283-98.
21. Nederveen, A.J., et al., *RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank.* Proteins, 2005. **59**(4): p. 662-72.
22. Ortega Roldan, J.L., et al., *The high resolution NMR structure of the third SH3 domain of CD2AP.* J Biomol NMR, 2007. **39**(4): p. 331-6.
23. Dolinsky, T.J., et al., *PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W665-7.
24. Hooft, R.W., C. Sander, and G. Vriend, *Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures.* Proteins, 1996. **26**(4): p. 363-76.
25. Wilkins, D.K., et al., *Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques.* Biochemistry, 1999. **38**(50): p. 16424-31.
26. Varela, L., et al., *A single mutation in an SH3 domain increases amyloid aggregation by accelerating nucleation, but not by destabilizing thermodynamically the native state.* FEBS Lett, 2009. **583**(4): p. 801-6.
27. Morel, B., S. Casares, and F. Conejero-Lara, *A single mutation induces amyloid aggregation in the alpha-spectrin SH3 domain: analysis of the early stages of fibril formation.* J Mol Biol, 2006. **356**(2): p. 453-68.
28. Dosset, P., et al., *Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data.* J Biomol NMR, 2000. **16**(1): p. 23-8.
29. Larson, S.M. and A.R. Davidson, *The identification of conserved interactions within the SH3 domain by alignment of sequences and structures.* Protein Sci, 2000. **9**(11): p. 2170-80.
30. Guerois, R., J.E. Nielsen, and L. Serrano, *Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations.* J Mol Biol, 2002. **320**(2): p. 369-87.

31. Sadqi, M., et al., *The native state conformational ensemble of the SH3 domain from alpha-spectrin.* Biochemistry, 1999. **38**(28): p. 8899-906.
32. Jarymowycz, V.A. and M.J. Stone, *Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences.* Chem Rev, 2006. **106**(5): p. 1624-71.
33. Ferreon, J.C., et al., *Solution structure, dynamics, and thermodynamics of the native state ensemble of the Sem-5 C-terminal SH3 domain.* Biochemistry, 2003. **42**(19): p. 5582-91.
34. Wang, C., N.H. Pawley, and L.K. Nicholson, *The role of backbone motions in ligand binding to the c-Src SH3 domain.* J Mol Biol, 2001. **313**(4): p. 873-87.
35. Horita, D.A., et al., *Dynamics of the Hck-SH3 domain: comparison of experiment with multiple molecular dynamics simulations.* Protein Sci, 2000. **9**(1): p. 95-103.
36. Knapp, S., et al., *Thermal unfolding of small proteins with SH3 domain folding pattern.* Proteins, 1998. **31**(3): p. 309-19.
37. Luque, I. and E. Freire, *Structure-based prediction of binding affinities and molecular design of peptide ligands.* Methods Enzymol, 1998. **295**: p. 100-27.
38. Chen, S., et al., *Abl N-terminal cap stabilization of SH3 domain dynamics.* Biochemistry, 2008. **47**(21): p. 5795-803.
39. Wuthrich, K., *NMR of proteins and nucleic acids*, ed. J.W. Sons. 1986.
40. Hwang, T.L. and A.J. Shaka, *Multiple-pulse mixing sequences that selectively enhance chemical exchange or cross-relaxation peaks in high-resolution NMR spectra.* J Magn Reson, 1998. **135**(2): p. 280-7.
41. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.
42. Johnson, B.A., *Using NMRView to visualize and analyze the NMR spectra of macromolecules.* Methods Mol Biol, 2004. **278**: p. 313-52.
43. Goddart, T.D.a.K., D.G., University of California, 1993.
44. Ruckert, M. and G. Otting, *Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments.* Journal of the American Chemical Society, 2000. **122**(32): p. 7793-7797.
45. Brutscher, B., *Accurate measurement of small spin-spin couplings in partially aligned molecules using a novel J-mismatch compensated spin-state-selection filter.* J Magn Reson, 2001. **151**(2): p. 332-8.
46. Lescop, E., P. Schanda, and B. Brutscher, *A set of BEST triple-resonance experiments for time-optimized protein resonance assignment.* J Magn Reson, 2007. **187**(1): p. 163-9.

47. Schanda, P., H. Van Melckebeke, and B. Brutscher, *Speeding up three-dimensional protein NMR experiments to a few minutes.* J Am Chem Soc, 2006. **128**(28): p. 9042-3.
48. Ortega-Roldan, J.L., et al., *Accurate characterization of weak macromolecular interactions by titration of NMR residual dipolar couplings: application to the CD2AP SH3-C:ubiquitin complex.* Nucleic Acids Res, 2009. **37**(9): p. e70.
49. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 5): p. 905-21.
50. Herrmann, T., P. Guntert, and K. Wuthrich, *Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS.* J Biomol NMR, 2002. **24**(3): p. 171-89.
51. Schanda, P., E. Kupce, and B. Brutscher, *SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds.* J Biomol NMR, 2005. **33**(4): p. 199-211.
52. Schanda, P. and B. Brutscher, *Hadamard frequency-encoded SOFAST-HMQC for ultrafast two-dimensional protein NMR.* J Magn Reson, 2006. **178**(2): p. 334-9.
53. Bai, Y., et al., *Primary structure effects on peptide group hydrogen exchange.* Proteins, 1993. **17**(1): p. 75-86.
54. Farrow, N.A., et al., *Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation.* Biochemistry, 1994. **33**(19): p. 5984-6003.

Chapter 4

Accurate characterization of weak macromolecular interactions by titration of NMR residual dipolar couplings: application to the CD2AP SH3-C:ubiquitin complex

Reprinted from Nucleic Acids Research, 2009, Vol.37, No. 9. (Jose Luis Ortega-Roldan, Malene Ringkjøbing-Jensen, Bernhard Brutscher, Ana I. Azuaga, Martin Blackledge and Nico A.J. van Nuland)

ABSTRACT

The description of the interactome represents one of key challenges remaining to structural biology. Physiologically important weak interactions, with dissociation constants above 100 mM, are remarkably common, but remain beyond the reach of most of structural biology. NMR spectroscopy, and in particular residual dipolar couplings (RDCs) provides crucial conformational constraints on intermolecular orientation in molecular complexes, but the combination of free and bound contributions to the measured RDC seriously complicates their exploitation for weakly interacting partners. We develop a robust approach for the determination of weak complexes, based on (1) differential isotopic labeling of the partner proteins facilitating RDC measurement in both partners, (2) measurement of RDC changes upon titration into different equilibrium mixtures of partially aligned free and complex forms of the proteins, (3) novel analytical approaches to determine the effective alignment in all equilibrium mixtures, (4) extraction of precise RDCs for bound forms of *both* partner proteins. The approach is demonstrated for the determination of the three dimensional structure of the weakly interacting CD2AP SH3-C:Ubiquitin complex (K_d=132±13mM) and is shown, using cross validation to be highly precise. We expect this methodology to extend the remarkable and unique ability of NMR to study weak protein-protein complexes.

INTRODUCTION

Following the successful development of structural genomic initiatives dedicated to the determination of the three dimensional conformation of a large number of proteins (1,2) attention is now turning to the characterization of the multitude of interactions between these proteins that control cellular processes and biological function (3-6). This paradigm, the description of the molecular basis of the interactome, is expected to provide a comprehensive portrayal of the overall interaction structure of an organism's proteome and thereby represents one of the major challenges for structural biology in the coming decade (7). Although very weak protein-protein interactions (dissociation constant K_d>10-4 M) are expected to be important for a vast range of cellular events, such as transcription and replication, signal transduction, transient formation of encounter complexes and assembly of protein complexes, they remain the least well characterized (8).

NMR is one of the most powerful tools for the study of biomolecular complexes, due to its sensitivity to protein-protein interactions with equilibrium dissociation constants varying over many orders of magnitude, including weak encounter complexes that can barely be detected using other biophysical techniques (8-10). In addition to the mapping of chemical shift

(CS) changes induced by the proximity of the partner protein, cross relaxation (NOe) derived intermolecular distance restraints and paramagnetic relaxation enhancements (9) residual dipolar couplings (RDCs) (11,12) have been shown to provide highly complementary orientational information that can be crucial for the determination of an accurate conformation of the complex (13-21). In the case of weak protein-protein complexes, where orientational constraints can be the most critical due to the insensitivity of NOESY experiments, the use of RDCs is seriously compromised by numerous experimental and theoretical complications. This is essentially because under conditions where the complex is too weak to be isolated experimentally, measured RDCs report on both bound and free forms of the molecule. Alignment characteristics of both free and bound forms of both proteins must therefore be determined and alignment levels accurately calibrated.

In this work we address these problems, and present a generally applicable protocol for the measurement, analysis and interpretation of RDCs for the refinement of the structure of weak protein-protein complexes. The protocol includes differential isotopic labeling of the two proteins (22) to allow the simultaneous measurement of RDCs at different molar ratios of the two partners, and uses a robust linear extrapolation approach to determine the bound form RDCs from both partner proteins from the same experimental mixtures. The protocol is shown via entirely independent cross-validation of data not used in the analysis to be highly accurate, and the importance of this methodology is clearly demonstrated by a detailed analysis of the significant structural errors that can be induced when residual components from the free forms of either protein contribute to the measured RDCs.

We apply this approach to the study of the complex formed between SH3-C, the third SH3 domain of CD2AP (CD2 associated protein) and ubiquitin. Ubiquitin is known to regulate a wide variety of cellular activities ranging from transcriptional regulation to cell signaling and membrane trafficking (23,24). Many cellular activities of ubiquitin are mediated by mono- rather than poly-ubiquitin, and its functions are deciphered by various ubiquitin-binding proteins. Similar to ubiquitin-binding domains, SH3 domains are found in proteins with different biological functions. SH3 domains form a highly conserved family of domains, but their amino acid composition varies at a few key sites, allowing for a wide range of molecular targets. Recently, it was found that a subset of SH3 domains constitutes a new, distinct type of ubiquitin-binding domains (25). The structure of the complex between Sla1 SH3-3 and ubiquitin shows that the ubiquitin binding surface of the Sla1 SH3 domain overlaps largely with the canonical binding surface for proline-rich ligands and that like many other ubiquitin-binding motifs, the SH3 domain engages the Ile44 hydrophobic patch of ubiquitin (26). Here we use NMR chemical shift perturbation and bound form RDCs for both proteins extrapolated from specifically developed RDC

titration experiments, to determine a structural model of the CD2AP SH3-C:Ubiquitin complex.

THEORETICAL ASPECTS

Residual dipolar couplings report on the orientation of internuclear vectors connecting two nuclei *i* and *j* with respect to the alignment tensor of a rigid molecule as follows:

$$D_{ij} = -\frac{\gamma_i \gamma_j \mu_0 h}{16\pi^3 r_{ij}^3}\left[A_a(3\cos^2\theta - 1) + \frac{3}{2}A_r \sin^2\theta\cos(2\varphi)\right] \quad (1)$$

A_a and A_r are the axial and rhombic components of the alignment tensors, and {q, φ} define the orientation of the internuclear vector with respect to the alignment tensor and r_{ij} is the internuclear distance. However, in the case of weak binding and fast exchange between free and bound forms of the proteins, experimental couplings, D_{ij}^{exp}, report on a combination of orientations of bound and free forms of the molecule (D_{ij}^{bound} and D_{ij}^{free})

$$D_{ij}^{exp} = p_{bound} D_{ij}^{bound} + (1 - p_{bound}) D_{ij}^{free} \quad (2)$$

assuming that the ratio of bound and free forms of the proteins is known, D_{ij}^{bound} can be determined using equation 2 from precise measurements of D_{ij}^{exp} and D_{ij}^{free}. Such a procedure has been applied to the study of weakly binding molecules to larger proteins where the nature of the larger system provides *a priori* knowledge allowing a simplification of the problem (18-21). Without knowledge of the structure of the complex the contribution of the measured values emanating from the free and bound forms of the proteins can be very difficult to quantitatively separate.

The use of RDCs to determine the relative orientation of partners in weak protein-protein complexes is also severely compromised by additional experimental and analytical difficulties, including the reproduction of identical absolute alignment conditions for the samples necessary for error-free subtraction of the free-form RDCs, and the associated potential for the propagation of experimental uncertainty that is inherent in the necessary subtraction required to derive values of D_{ij}^{bound}. RDCs from the bound form of the labelled proteins can of course be isolated under conditions where essentially all of these proteins are bound in the complex, but in the case of weak complexes this saturation limit requires potentially prohibitive concentrations of the partner protein.

We therefore propose a procedure that simultaneously determines the alignment characteristics of both free and bound forms of the proteins, and determines the level of alignment in each medium. Instead of measuring a single mixture, RDCs are measured over a range of titration mixtures m_i of free and bound forms of both proteins. In each of the mixtures the measured RDC, $D_{ij}^{m,\exp}$, is given by a slightly modified version of equation 2:

$$D_{ij}^{m,\exp} = \lambda_m D_{ij}^{m} = \lambda_m \left\{ p_{bound} D_{ij}^{bound} + (1 - p_{bound}) D_{ij}^{free} \right\} \quad (3)$$

where l_m is a scaling factor defined by the absolute level of alignment that is in turn determined by the concentration of the alignment medium, and p_{bound} spans different values for the two proteins. D_{ij}^{free} and D_{ij}^{bound} are given by:

$$D_{ij}^{free} = -\frac{\gamma_i \gamma_j \mu_0 h}{16\pi^3 r_{ij}^3}\left[A_a^{free}(3\cos^2\theta^{free} - 1) + \frac{3}{2} A_r^{free} \sin^2\theta^{free} \cos(2\varphi^{free}) \right] \quad (4)$$

$$D_{ij}^{bound} = -\frac{\gamma_i \gamma_j \mu_0 h}{16\pi^3 r_{ij}^3}\left[A_a^{bound}(3\cos^2\theta^{bound} - 1) + \frac{3}{2} A_r^{bound} \sin^2\theta^{bound} \cos(2\varphi^{bound}) \right] \quad (5)$$

A_a and A_r refer to the free or bound alignment tensors, and the two spherical coordinates {q,φ} refer to the orientation of the internuclear vector with respect to the alignment tensor in the two forms. Note that this simple linear relationship (equation 3) holds irrespective of changes in structure and dynamics of the site of interest in the complex. For some alignment media l_m can be estimated from the deuterium quadrupolar coupling from the D_2O present in the sample. This is not a precise metric however, and it is therefore preferable to exploit the combined dependence of equation 3 to determine the exact scaling factors, by adjusting l_m to maintain optimal linearity of D_{ij}^{m} relative to p_{bound}. The same scaling factor is applied to both proteins in the respective mixtures as differential isotope labeling allows simultaneous measurement of RDCs in the two proteins. All couplings from both proteins are used in this procedure ensuring a high level of precision and robustness implicit in this optimization. Following this adjustment the expected values of the bound forms of the RDCs from both proteins are determined from the resulting linear titration relationships established for each individual RDC. This procedure turns out to be highly robust and significantly more accurate than subtraction of a single value scaled on the basis of ^2D splitting (vide infra).

MATERIALS AND METHODS

Protein expression and purification

For clarity, we use the residue numbering of CD2AP SH3-C of PDB entry 2JTE throughout the text. Unlabelled, ^{15}N-labelled and ^{15}N,^{13}C-labelled CD2AP SH3-C was obtained as described (27). Unlabelled and ^{15}N-labelled ubiquitin was purchased from both Cortecnet and Spectra Stable Isotopes. ^{15}N/^{13}C-labelled ubiquitin was kindly provided by Varian Inc. Protein concentrations were determined by absorption measurements at 280 nm using an extinction coefficient of 13980 and 1450 $cm^{-1} \cdot M^{-1}$ for SH3-C and ubiquitin, respectively, determined using the ProtParam algorithm (www.expasy.ch).

NMR chemical shift perturbation

All NMR titration experiments were performed at 25ºC 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. CD2AP SH3-C and ubiquitin samples were prepared for NMR experiments in 93% H_2O/7% D_2O, 50 mM NaPi, 1 mM DTT at pH 6.0. The backbone amide and ^{15}N frequencies of CD2AP SH3-C under the above conditions, previously assigned at pH 2.0 (27), were obtained first by comparing 2D ^{15}N-HSQC spectra at pH 2.0, 3.0, 6.0 and 7.0 and confirmed by a single HNCACB triple resonance experiment acquired on ^{15}N,^{13}C-labelled CD2AP SH3-C at pH 6.0. A HNCACB triple resonance spectrum was also recorded on a ^{13}C/^{15}N-labelled ubiquitin to confirm backbone assignment at pH 6.0.

The SH3-binding site on ubiquitin was obtained by titrating with increasing amounts of unlabeled CD2AP SH3-C domain into a 0.25 mM ^{15}N-ubiquitin sample at pH 6.0, 25 ºC. Similarly, the ubiquitin-binding site on CD2AP SH3-C was obtained by titrating with increasing amounts of unlabeled ubiquitin into a 0.25 mM ^{15}N-SH3-C sample under the same conditions. The progress of the titrations was monitored by recording one-dimensional ^{1}H and two-dimensional ^{1}H-^{15}N HSQC spectra.

The magnitude of the chemical shift deviations (Dd) was calculated using the equation:

$$\Delta\delta_{ppm} = \sqrt{\left(\Delta\delta_{HN}\right)^2 + \left(\frac{\Delta\delta_N}{6.51}\right)^2} \qquad (6)$$

where the difference in chemical shift is that between the bound and free forms of the different proteins.

All NMR data were processed using NMRPipe (28) and analyzed by NMRView (29).

Measurement of RDCs

RDCs were measured in samples partially aligned in a liquid-crystalline medium consisting of a mixture of 5% penta-ethyleneglycol monododecyl ether ($C_{12}E_5$) and hexanol (30). For the $^{13}C/^{15}N$-labelled protein (SH3) in the free form or in diverse mixtures of free and complex, a set of 4 different RDCs ($^1D_{NH}$, $^1D_{C_\alpha C'}$, $^2D_{HN\,C'}$ and $^1D_{C_\alpha H_\alpha}$) was measured per sample using 3D BEST-type HNCO or HNCOCA experiments (31,32). Coupling constants were obtained from line splittings in the ^{13}C dimension using the nmrPipe nlinLS fitting routine or using Sparky (33). For the ^{15}N-labelled protein (ubiquitin) in free or in the complex, $^1D_{NH}$ were measured from a pair of spin-state-selected ^1H-^{15}N correlation spectra recorded using the pulse sequence shown in figure S4. The pulse sequence uses a DIPSAP filter (34) for J-mismatch compensated spin-state selection, and the BEST concept for longitudinal relaxation and sensitivity enhancement (31,32). In addition, signals from the doubly ($^{13}C/^{15}N$) labeled binding partner are removed by additional transfer steps from ^{15}N to ^{13}CO. After this transfer step the presence of orthogonal coherences for the spin systems from ^{15}N-only and $^{13}C/^{15}N$ labeled proteins is exploited to suppress the unwanted signals by means of pulsed field gradients and phase cycling. Total measurement time for one titration point is approximately 1 day.

RDC Refinement of the SH3-C domain of CD2AP in complex with ubiquitin

The program SCULPTOR (35) has recently been developed as an addition to the program CNS (36). The refinement protocol involves a restrained MD calculation using the standard CNS force field. Starting structures were taken from a selection of 10 structures determined using HADDOCK (37) based on ambiguous intermolecular restraints (AIRs). Initial sampling was further increased by including a sampling period of 10ps at 700K that allows for the SH3 domain to reorient freely without RDCs or AIR restraints. Following this both ubiquitin and SH3 conformations are fixed and the alignment tensor is allowed to evolve freely to determine initial estimates from the SH3 structure (38,39) (there are four RDC types available for this structure). Both molecules and tensors are then freed with the initial sampling period at 1000K for 5ps during which time both AIRS and RDCs are scaled from 0.1% to 100% of their final values. The backbone conformation of the segment (1-70) of ubiquitin is restrained to its initial coordinates using a harmonic potential, while the experimentally measured NOes from the free form of SH3 are used as restraints. Note the

RDCs measured in the bound form can be used to determine structural changes upon binding, as the RDC titration approach provides accurate constraints from bound forms of both partners irrespective of differences in structure and dynamics between the free and bound forms (see below). AIR restraints are used as described in the supporting information. Sampling of restraints is followed by a 5ps sampling stage and slow cooling over 5ps to 100K and energy minimization. The protocol is repeated 30 times for each starting structure.

RESULTS AND DISCUSSION

Chemical Shift Perturbation of Ubiquitin and SH3

NMR chemical shifts of protein residues are highly sensitive to changes in the local environment, and chemical shift perturbations are therefore widely used to map intermolecular interfaces of protein complexes (40) and to drive the docking of two interacting partners in order to obtain a structural model of the complex (41,42). Titrations of ^{15}N-labeled ubiquitin with increasing amounts of SH3-C domain at 25 ºC and vice versa caused a selective shift of amide proton and nitrogen resonances of several ubiquitin and SH3-C residues (figure 1), indicating a specific union between the two proteins. The most significant changes in the chemical shift of the HSQC cross-peaks of the SH3 domain can be observed at the *RT loop* (residues 18-23), the *nSrc loop* (residues 37-39), the beginning of the b-*III strand* (residues 42-44), and residues 54 to 58. Considerable changes in the chemical shift of the HSQC cross-peaks of ubiquitin are mainly observed in two typical binding regions, namely that of Ile44 and of Gly76.

The SH3-Ubiquitin complex is in fast exchange with the free forms of the partners as shown from the chemical shift titration in figures 1A and 1B (upper panels). Thus, no significant line broadening of the resonances is observed during the titration. On the basis of a simultaneous fit of the combined $^{1}H^{N}$ and ^{15}N shifts of residues 18, 19, 20, 21, 38, and 44 (SH3) to titration of added ubiquitin, and 42, 45, 47, 68 and 73 (Ubiquitin) to titration of added SH3 a K_d of 132±13 mM was determined (figure 1C, D). The affected sites on SH3 are identified from the extent of the chemical shift perturbations upon addition of ubiquitin and vice versa (figure 1A and 1B, lower panels), and these sites were transformed into ambiguous intermolecular distance restraints (AIRs, see Methods section in supporting information). The program HADDOCK (37) was used to calculate the structure of the complex on the basis of these restraints. The 200 lowest energy structures were refined in a shell of water, and one representative structure of each of 10 identified clusters were selected (see supporting information figure S1) for further refinement of the conformation of the complex using RDCs measured of both proteins in the bound forms.

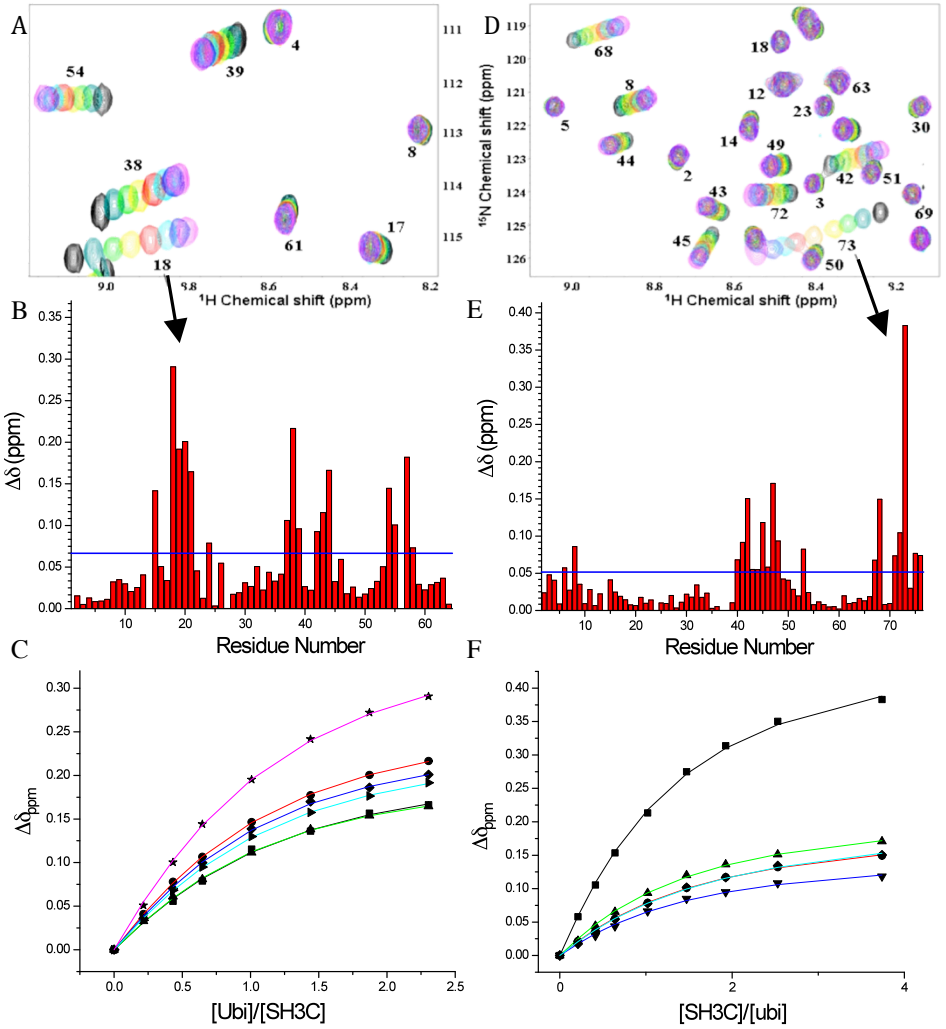


Figure 1. Monitoring the binding between ubiquitin and CD2AP SH3-C by NMR chemical shift perturbations. A. *^{1}H and ^{15}N chemical shift changes upon titrating ubiquitin into a ^{15}N-labelled CD2AP SH3-C solution. Region of ^{1}H-^{15}N HSQC spectra recorded at increasing amounts of ubiquitin (black to magenta). The chemical shift changes (Δδ) between the first and last titration point (equation 5) are presented in* ***B. D and E.*** *^{1}H and ^{15}N chemical shift changes upon titrating SH3-C into a ^{15}N-labelled ubiquitin solution. Blue solid lines indicate the mean chemical shift changes that were used to define the active SH3 and ubiquitin residues used in HADDOCK.* ***C.*** *Changes in chemical shift, Δδ, of selected residues in ^{15}N-labeled SH3 upon addition of unlabeled ubiquitin: black Trp44; red Thr38 ; green Asp21 ; blue Glu20 ; cyan Asn19 ; magenta Thr18.* ***F.*** *Changes in chemical shift, Δδ, of selected residues in ^{15}N-labeled ubiquitin upon addition of unlabeled CD2AP SH3-C: black Leu73; red His68 ; green Gly47 ; blue Phe45 ; cyan Arg42.*

Titration of RDCs for Ubiquitin and SH3 in the Protein-Protein Complex

SH3 was uniformly isotopically labelled ^{13}C and ^{15}N, and ubiquitin was uniformly ^{15}N labelled. This allowed the use of isotopically filtered RDC measurements (see Materials and Methods) such that couplings from both partners were measured in three mixtures (m_1, m_2 and m_3) of the two components. ^{15}N-$^{1}H_N$ RDCs were measured in ubiquitin and ^{15}N-$^{1}H_N$, $^{13}C'$-$^{1}H^N$, $^{13}C'$-$^{13}C_\alpha$ and $^{13}C_\alpha$-$^{1}H_\alpha$ were measured for SH3. This labelling scheme was adopted to simplify spectral analysis when RDCs are measured from both partners in the same sample, an integral part of the method developed here (22). These RDCs were also measured in independently aligned free forms of both molecules (samples $m_{0,ubi}$ and $m_{0,sh3}$). For each partially aligned sample, p_{bound} was determined from the ratios of intensities (volumes) of resonances in ^{1}H-^{15}N HSQC spectra containing peaks from both SH3 and ubiquitin. The mixtures (m_1, m_2 and m_3) were thus estimated to represent p_{bound} values of (0.44, 0.75 and 0.84) for SH3 and (0.79, 0.61 and 0.49) for ubiquitin.

Correlation plots of $^{1}D_{NH}$ measured in the free forms of SH3 and ubiquitin compared to values measured at the other three mixtures are indicative of differently oriented alignment tensors in the free and bound forms of both proteins, in particular SH3 (figure 2 and table 1). RDCs measured in free forms and mixtures of SH3 and ubiquitin were compared to known structures of the two proteins (pdb codes: 1D3Z (43) for ubiquitin and the RDC refined structure of SH3) using the program MODULE (44). All data sets fit reasonably well to these structures, as illustrated in figures S2 and S3 in supporting information and table 1.

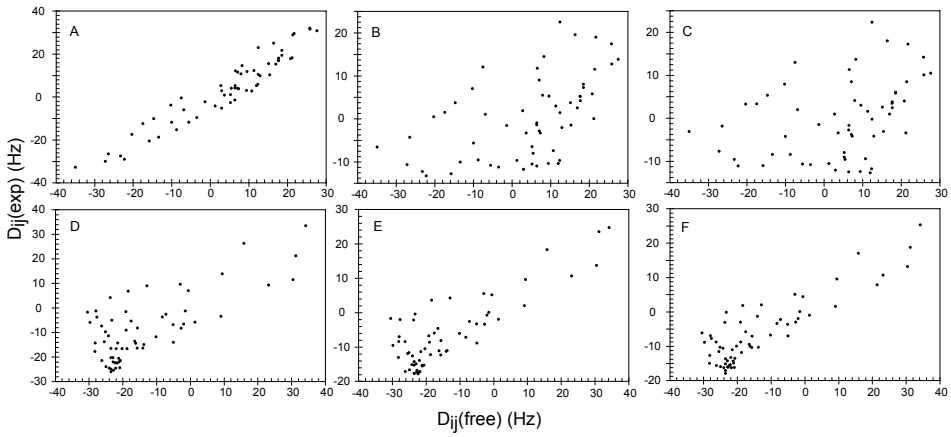


Figure 2. Correlation plots of experimental ^{15}N-$^{1}H^N$ RDC data sets. (A-C) *RDCs in the free form of SH3 versus the RDCs in mixture m_1 (A), m_2 (B) and m_3 (C).* ***(D-F).*** *RDCs in the free form of ubiquitin versus RDCs in mixture m_1 (D), m_2 (E) and m_3 (F).*

In the case of the RDCs emanating from mixtures of D_{ij}^{bound} and D_{ij}^{free}, the effective alignment tensor results from a linear combination of the free and bound alignment tensors. The effective tensors resulting from these different combinations can vary significantly in terms of rhombicity and orientation (simulated data shown in table S2 in supporting information). It is important to note that the fit of the known structure to the RDCs from any mixture will be equally good, so that the presence of a contribution from the unbound form cannot be identified on the basis of the data reproduction, but as we will see below, these effects can strongly influence the final conformation of the complex. It is therefore important to accurately estimate the true values of the bound RDCs. The procedures we have developed to achieve this are presented in the Theoretical section above and the results described below.

Table 1. Alignment tensors of different experimental mixtures. *[a]Tensor eigenvalues (scaled by appropriate l value) resulting from fit of experimental RDCs measured from free Ubiquitin to the structure 1d3z. [b]Tensor eigenvalues resulting from fit of fully bound Ubiquitin RDCs. [c]Simultaneous refinement of the structure of 1d3z and alignment tensor eigenvalues using RDCs from fully bound Ubiquitin. [d]Simultaneous refinement of the complex structure of ubiquitin/SH3 complex using RDCs from fully bound forms of Ubiquitin and SH3. [e]Simultaneous refinement of the structure of SH3 and alignment tensor eigenvalues using RDCs from fully bound form of SH3. [f]Fit of experimental RDCs from bound form of SH3 to RDC-refined structure of free-form SH3. [g]Simultaneous refinement of SH3 (pdb code 2JTE) and alignment tensor eigenvalues using RDCs from SH3 alone.*

	Free [a] Ubi	m_3	m_2	m_1	**Bound [b] Ubi**	**Bound [c] Ubi ref**
$A_a (10^{-4})$	12.0	12.2	12.6	13.2	13.3	13.7 ± 0.2
$A_r (10^{-4})$	1.10	2.1	2.4	4.5	4.5	4.3 ± 0.1
α (°)	124	-24	-29	-37	-40	-
β (°)	148	38	39	45	47	-
γ (°)	173	7.4	10	20	22	-

	SH3/Ubi [d] ref	**Bound [e] SH3 ref**	**Bound [f] SH3**	m_3	m_2	m_1	**Free [g] SH3**
$A_a (10^{-4})$	12.9 ± 0.1	12.4 ± 0.1	11.8	10.0	10.0	-14.7	-25.3
$A_r (10^{-4})$	3.7 ±0.1	3.0 ± 0.1	3.0	0.3	1.2	-6.7	-7.3
α (°)	-	-	-126	-126	-144	-82	-107
β (°)	-	-	54	116	108	150	148
γ (°)	-	-	87	-99	-104	-5	-28

Extraction of Bound Form RDCs for Ubiquitin and SH3

A total of 1168 experimentally measured RDCs are used to determine the four global scaling parameters l_i and the bound-form values of *both* proteins are extrapolated from the linear build-up for each individual RDC. This approach involves fixing the scaling factor (l_3) for one of the three mixtures (m_3) of measured couplings, while RDCs from the remaining mixtures are used to simultaneously determine optimal values for l_1, l_2, $l_{0,ubi}$ and $l_{0,sh3}$. The scaling factors are applied to data measured on both proteins simultaneously such that while l_1 corresponds to the required scaling for the equilibrium with majority free form of SH3, it also corresponds to the required scaling for the equilibrium with majority of bound form ubiquitin. This adjustment is independent of which scaling factor is initially fixed. Comparison of the optimal scaling parameters extracted using this approach with the ratio of the overall alignment as estimated from the ^{2}H splitting indicates very good similarity between the independently estimated values (table 2). The approach can therefore be applied with confidence for alignment media or experimental conditions for which the ^{2}H splitting does not reflect a linear measure of the effective protein alignment. We note that in cases where the level of alignment is accurately known, an analogous procedure can be applied to adjust the molar ratio of the components for each mixture.

Table 2. Comparison of D_2O scaling parameters with those extracted from fits. *[a]number in parentheses indicated the scaled splitting relative to m_3 mixture, for comparison with λ_m values scaled to λ_3*

Sample	m_1	m_2	m_3	$m_{0,sh3}$	$m_{0,ubi}$
Concentration of SH3 (mM)	0.90	0.83	0.78	1.00	0.00
Concentration of ubiquitin (mM)	0.50	1.03	1.33	0.00	0.50
p_{bound} for SH3	0.44	0.75	0.84	0.00	-
p_{bound} for ubiquitin	0.79	0.61	0.49	-	0.00
Experimental D_2O splittings (Hz) [a]	29 (1.16)	25 (1.00)	25 (≡ 1)	18 (0.72)	37 (1.48)
λ_m	1.233	1.027	≡ 1	0.643	1.643

Figure 3 shows the dependence of arbitrarily selected couplings on p_{bound} for both SH3 and ubiquitin after optimization of l_1, l_2, $l_{0,ubi}$ and $l_{0,sh3}$ relative to l_3. Linearity of RDCs with respect to p_{bound} is very clearly observed throughout the molecule. This linearity is expected to exist irrespective of the local structure and dynamics experienced in the two forms, as long as

the exchange can be described in terms of a two state system and that it is fast on the chemical shift and dipolar coupling time scale. To further demonstrate the available precision of this analysis, one tenth of all RDCs were randomly removed from the total data set and retained for comparison with predicted values derived from the remaining RDCs. Comparisons of experimental and predicted data agree within experimental error for all RDCs (figure 4). This figure also illustrates the relative precision of the different coupling types relative to their experimental range, with slightly higher dispersion for the smaller $^{13}C'$-$^{1}H^{N}$ and $^{13}C'$-$^{13}C^{a}$ couplings. This analysis can be used to estimate the precision of the extrapolated bound-form couplings in the range of 0.5-1.0 Hz, that combines all contributions to uncertainty, including sample preparation, experimental uncertainty, analytical extraction of the RDCs and extrapolation of the bound-form values and estimation of the fraction of free and bound forms.

Refinement of the Complex using Bound Form RDCs for Ubiquitin and SH3

Following this procedure the eigenvalues of the alignment tensors of the scaled free and bound forms represent expected values for a common level of alignment for the two proteins. Comparison of the extrapolated bound form RDCs (open circles in figure 3) with known structures of the free form proteins (table 1) shows that the optimal bound form tensor eigenvalues are slightly smaller for SH3 than for Ubiquitin (A_a≈13.3 10^{-4}, A_r≈4.5 10^{-4}, compared to 11.8 10^{-4} and 3.8 10^{-4}). This comparison may be biased by a dependence both on angular sampling of RDC-relevant vectors in the two proteins and on the quality of the structures used to determine the tensors (45,46) (a high resolution RDC-refined structure of ubiquitin – 1D3Z and the free form structure of SH3 refined with the four types of RDCs measured on the aligned free form). We therefore carried out a control calculation, wherein the two structures were refined against independent tensors. This leads to optimal tensors for the two refined protein structures of A_a=13.7 10^{-4}, A_r=4.3 10^{-4} for ubiquitin and A_a=12.3 10^{-4}, A_r=3.0 10^{-4} for SH3 (table 1). The apparent differences in eigenvalues of both refined structures are therefore small, and we now focus on the determination of the average orientation of the two proteins in the complex.

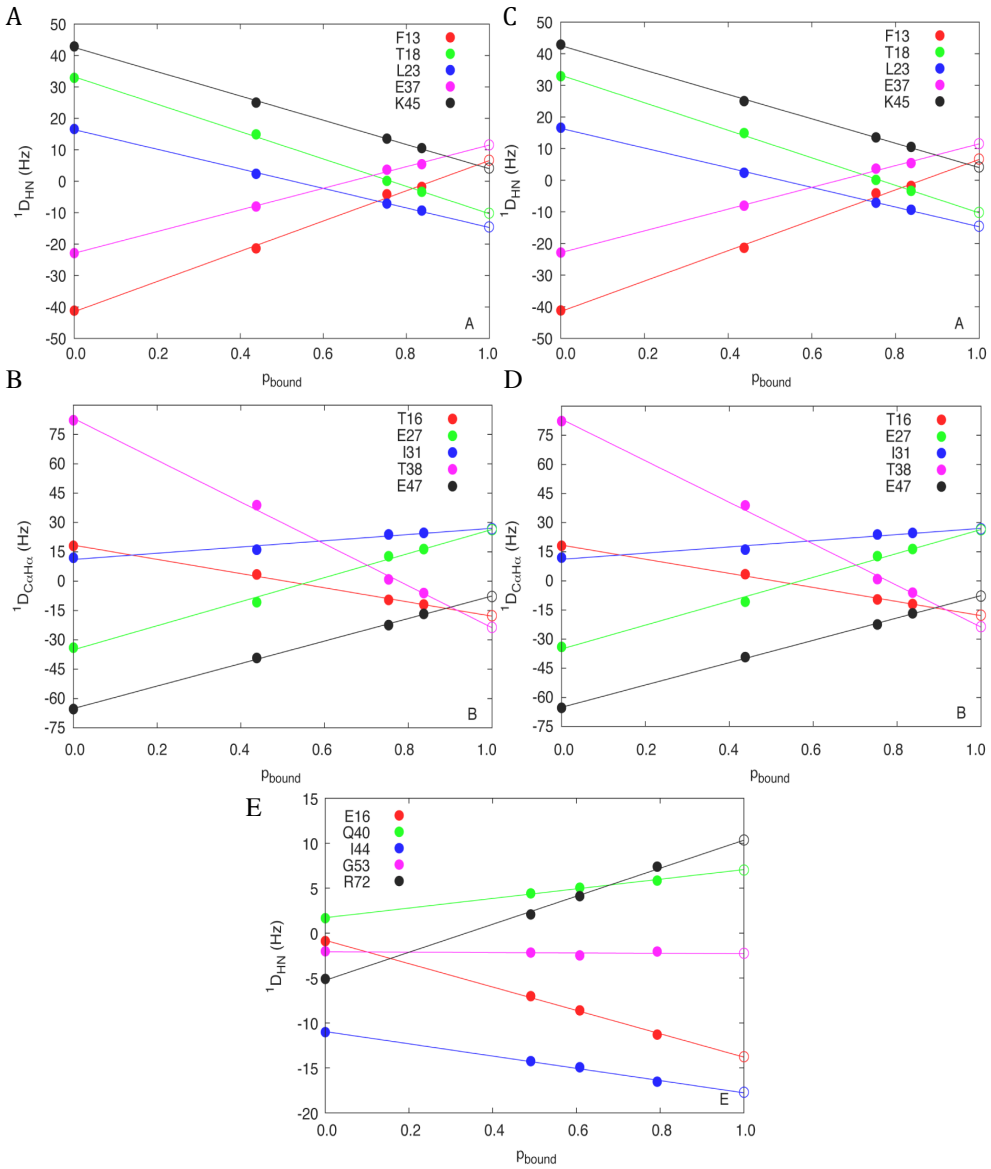


Figure 3. Plots of linearly changing RDCs with respect to population of the bound state (x-axis). *The fully bound values (open circles) represent the values determined by optimization of equations (2). The remaining values are experimental values, adjusted by optimizing the linearity of the slope by scaling the four parameters l_1, l_2, $l_{0,ubi}$ and $l_{0,sh3}$ relative to l_3. Five sites are shown for each RDC type (A – D SH3, E Ubiquitin). The selected sites show fits of average quality.*

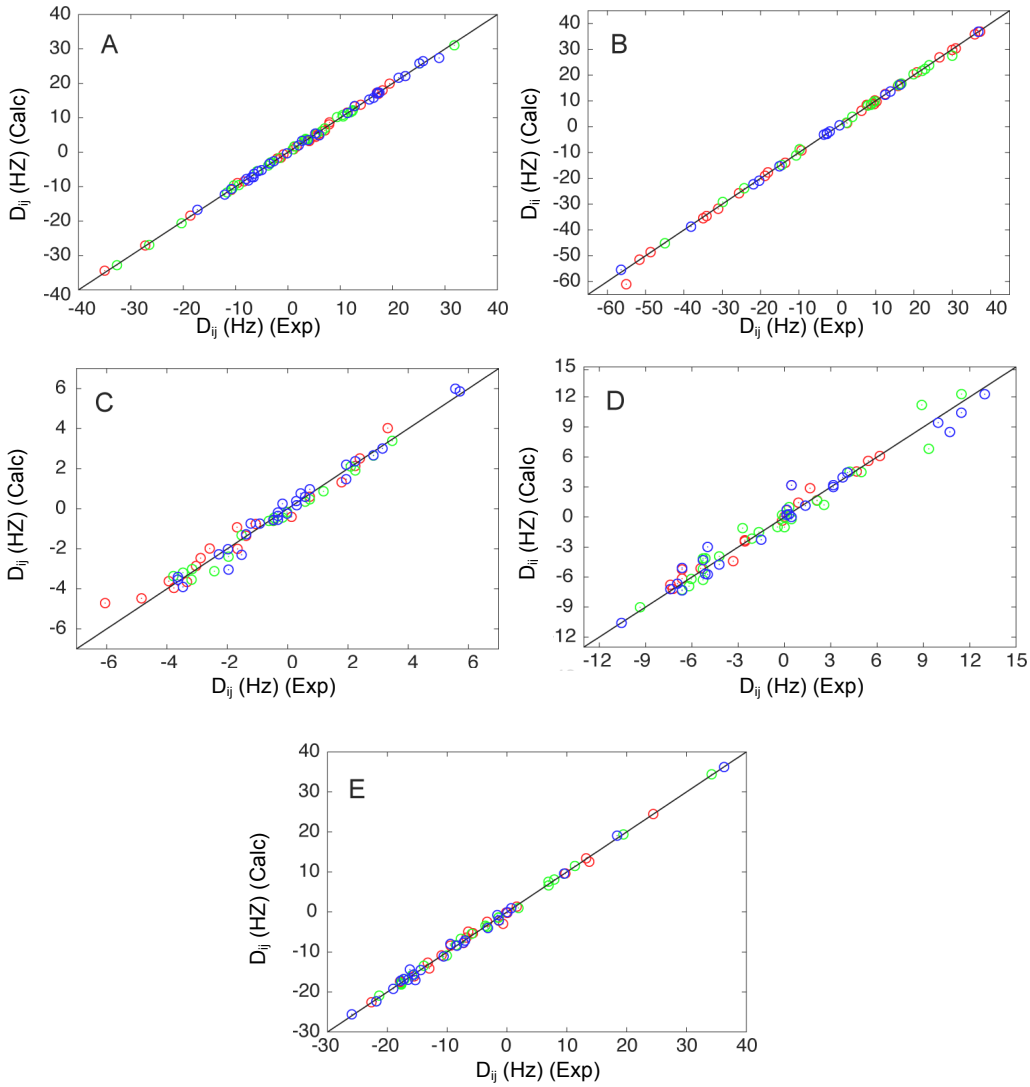


Figure 4. Cross-validation of RDC titration analysis. *(A-E) Back calculated points from linear fitting approach shown in figure 3. 120 different RDCs were removed from the analysis, back-calculated (y-axis) and compared to the experimental value (x-axis). The results of three such calculations are shown in each case (each colour represents one of the three analyses) (A) SH3 ^{15}N-$^{1}H_N$ RDCs, (B) SH3 $^{13}C_\alpha$-$^{1}H_\alpha$ RDCs, (C) SH3 $^{13}C_\alpha$-$^{13}C'$ RDCs, (D) SH3 $^{13}C'$-$^{1}H_N$ RDCs, (E) Ubiquitin ^{15}N-$^{1}H_N$ RDCs.*

Bound-form RDCs determined in this way for both proteins were introduced into a structure refinement procedure starting from conformers from different clusters sampled by HADDOCK using only chemical shift perturbation as intermolecular restraint. The refinement was carried out using the program Sculptor, recently interfaced into CNS, using a refinement protocol assuming a common alignment tensor for both domains whose components are simultaneously determined during the structure refinement (see Methods in supporting information). The structures with the lowest target function (combining RDC, AIR and NOe violations) in the ensemble are shown in figure 5A and statistical analysis summarized in table S3.

The resulting structure of the complex is importantly different from the Sla1 SH3-3:Ubiquitin, (25,26) and the CIN85 SH3-C:Ubiquitin structure (47). A key affinity and specificity determinant for ubiquitin-binding was appointed to Phe409 of Sla1, located at the heart of the hydrophobic interface in the SH3-ubiquitin complex. Other SH3 domains with a tyrosine at this position were found to be incompetent to bind ubiquitin (25). Moreover, the Phe409 to Tyr mutation in Sla1 was shown to abolish ubiquitin binding confirming the key role of the phenylalanine residue at this position (25). In our structure, the corresponding phenylalanine residue in the CD2AP SH3-C:Ubiquitin complex is placed at the edge of the binding interface (figure 5B). In contrast to Sla1 SH3-3, the replacement of this phenylalanine in CD2AP SH3-C by a tyrosine does not abolish ubiquitin binding as monitored by fluorescence, ITC and NMR experiments (data not shown), thus confirming the correctness of our model and indicating that in this case the phenylalanine does not play a key role for the affinity and specificity of ubiquitin binding. The intriguing question that remains is which residues and/or which region of a particular ubiquitin-binding SH3 domain determines the mode/orientation of binding to ubiquitin and if this difference in binding mode is somehow related to a difference in function. Further progress in this direction will require a more detailed structural and functional study of these complexes, but we note that the recent observation of alternative binding modes present in the formation of this complex already provides evidence for transient encounter complexes that are sampled during the course of this interaction (48).

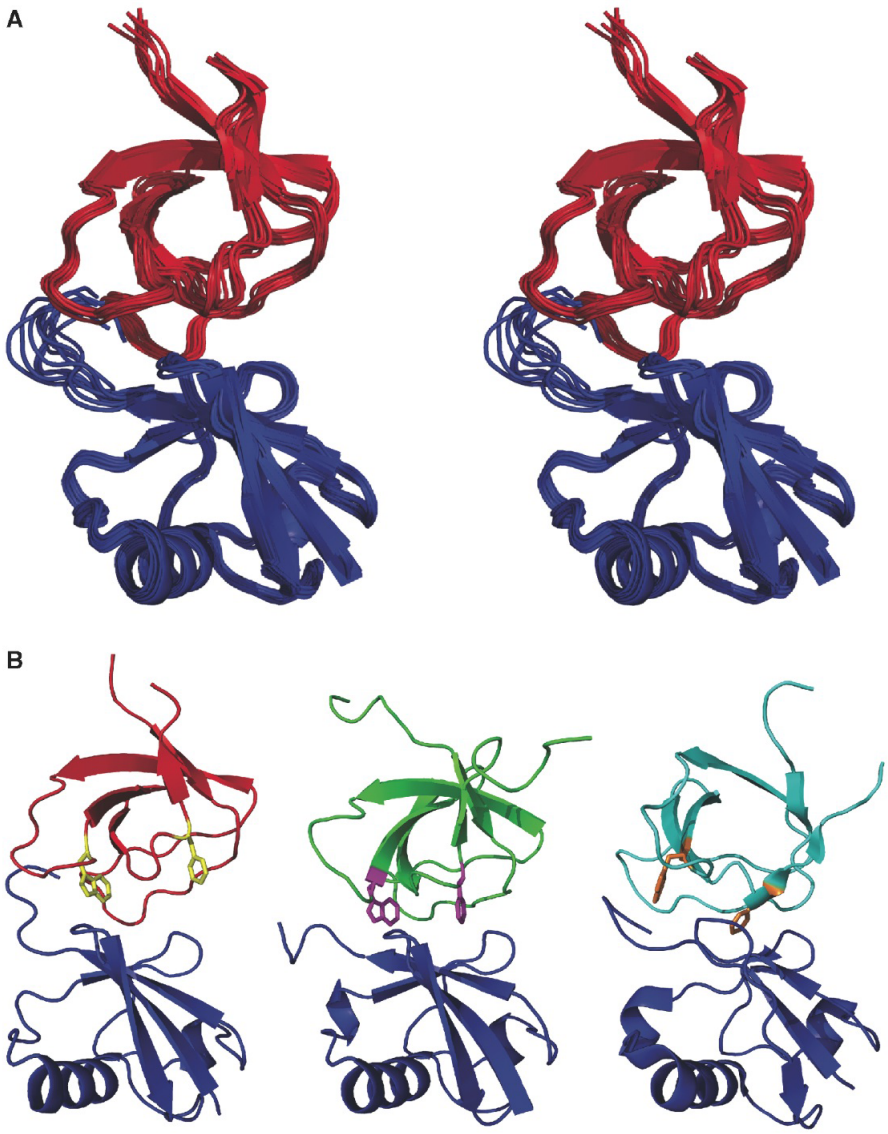


Figure 5. Structural model of the complex between CD2AP SH3-C and ubiquitin. ***A.*** *Stereo representation of the ensemble of 10 lowest-energy structures derived from the RDC titration protocol (SH3-C in red, ubiquitin in blue).* ***B.*** *Comparison between the CD2AP SH3-C:Ubiquitin (SH3-C in red, ubiquitin in blue), Sla1 SH3-3:Ubiquitin complex (PDB entry 2JTA; SH3-C in green, ubiquitin in blue) and CIN85 SH3-C:Ubiquitin (PDB entry 2K6D; SH3 in cyan, ubiquitin in blue). Trp43 and Phe59 in CD2AP SH3-C are shown in yellow sticks and the equivalent residues in Sla1 SH3-3 and CIN85 SH3-C are shown in magenta and orange sticks, respectively. The SH3:Ubiquitin complexes were superimposed on the backbone atoms of residues 4-71 of ubiquitin (RMSD of 0.74 and 1.32 Å for Sla1 and CIN85 to CD2AP, respectively).*

Comparison to a determination of RDCs from weak complexes with standard protocols.

The true values of the bound form RDCs are not known *a priori*, precluding direct assessment of the relative accuracy of this approach compared to a classical extrapolation from a single mixture. Nevertheless we are able to compare experimentally measured values from one of the mixtures. All ^{15}N-$^{1}H^{N}$ RDCs measured in SH3 were therefore removed from mixture m_3, the analysis was repeated and experimental and predicted values were compared (figure S6B in supporting information). We have compared this analysis to a prediction of SH3 RDCs from mixture m_3 on the basis of values measured in m_0 and m_1 using a standard extrapolation:

$$D_{ij}^{m_3} = (\lambda_1 p_3 D_{ij}^{m_1} - \lambda_{0,sh3}(p_3 - p_1) D_{ij}^{\text{free}}) / \lambda_3 p_1 \qquad (7)$$

The populations p are assumed to be accurately known in both cases and in this case the scaling terms l represent the experimentally measured D_2O splitting (table 2). This case is equivalent to a single mixture where over 50% of SH3 molecules are in the complex. As shown in figure S6 the ability to reproduce experimental RDCs is significantly improved using the RDC titration approach. The two-point approach, that is directly analogous to standard extrapolation methods based on measurement of a single mixture, suffers from two drawbacks: the combination of experimental errors of the two experimental points that results in random fluctuations of the predicted RDCs and the uncertainty in the actual alignment level that scales all couplings by an unknown systematic error. The combination of these errors results in a poor reproduction of experimental data, and indicates that this kind of approach is likely to provide highly inaccurate RDCs for refinement of molecular orientation. Neither of these sources of error can be corrected without further measurements, and both sources are minimized and at least partially corrected in the RDC titration approach. The possible consequences of these errors in terms of molecular structure are discussed below.

Refinement against RDCs measured in non-saturated conditions.

The real importance of determining accurate alignment tensors for the precise determination of weakly interacting proteins is illustrated in figure 6, where data were simulated from the experimentally determined alignment tensors of free and complexed SH3 and Ubiquitin, and used as restraints to orient the partner proteins. Three conditions of incomplete saturation were simulated: 90% saturation for both proteins, 80% saturation and 70% saturation. In each case the remainder of the protein was assumed to be in its free form.

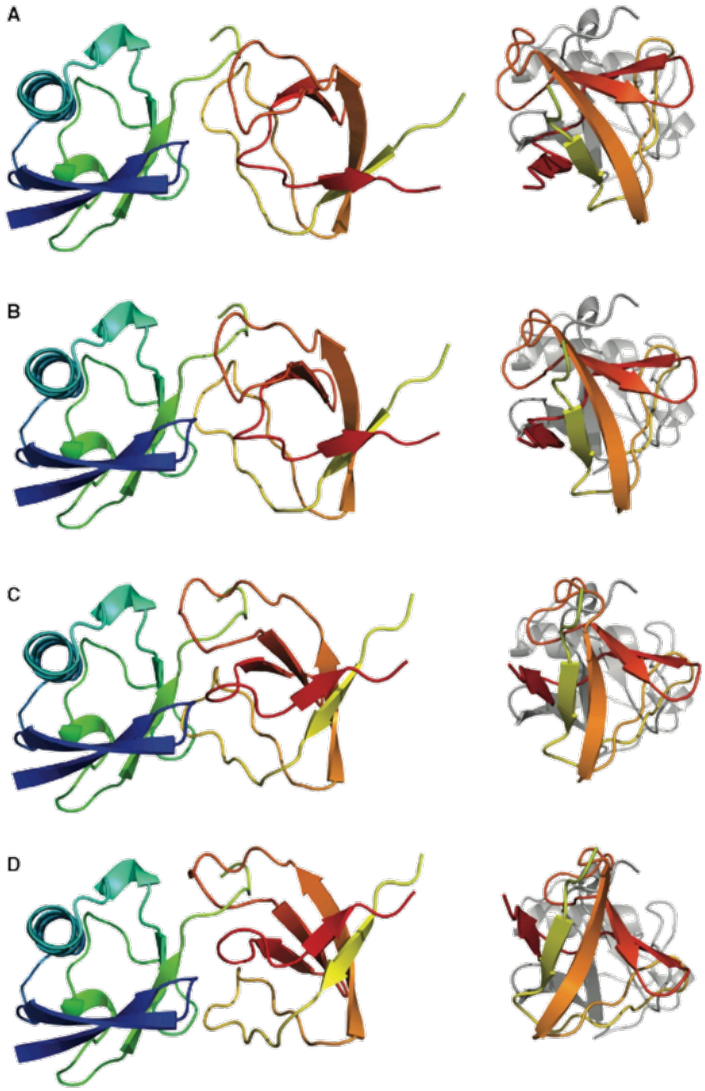


Figure 6. Effect of non-saturation on effective alignment and orientation of the SH3-Ubiquitin complex. *Data were simulated from the bound and free states of ubiquitin and SH3 and mixed in ratios of 1.00:0.0 (A), 0.9:0.1 (B), 0.8:0.2 (C) and 0.7:0.3 (D). The relative orientation was determined by aligning the axes of the effective alignment tensors for the two domains. Right hand panels represent a rotation of 90° about the vertical axis with respect to the left-hand panels The ubiquitin domain is held fixed for the purposes of comparison in the figure. The alignment tensors for SH3 and ubiquitin are respectively given by (A)* A_a*=12.2* 10^{-4}*,* A_r*=3.2* 10^{-4}*;* A_a*=13.3* 10^{-4}*,* A_r*=4.3* 10^{-4}*;(B)* A_a*=10.9* 10^{-4}*,* A_r*=2.3* 10^{-4}*;* A_a*=12.8* 10^{-4}*,* A_r*=4.0* 10^{-4}*; (C)* A_a*=10.0* 10^{-4}*,* A_r*=2.0* 10^{-4}*;* A_a*=12.4* 10^{-4}*,* A_r*=3.7* 10^{-4}*; (D)* A_a*=9.6* 10^{-4}*,* A_r*=3.0* 10^{-4}*;* A_a*=12.0* 10^{-4}*,* A_r*=3.4* 10^{-4}*. Differences in orientation relative to the fully bound form (A) can be measured by the differences in the Euler rotations of the alignment tensor axes for the two molecules: (B) da=10°, db = 10°, dg = 2°; (C) da=32°, db = 20°, dg = 7°; (D) da=54°, db = 29°, dg = 12°.*

The optimal configurations of the complexes, determined after orientation of the partner proteins such that their effective alignment tensor axes were coaxial and intermolecular contacts best reproduced, are shown in figure 6. The orientation of the partners very clearly changes, and is evidently incorrect even for complexes refined with data emanating from samples containing relatively high populations of complexed relative to free form. When the dissociation constant is sufficiently weak, it is therefore extremely risky to exploit RDC data sets from a single mixture even when the data are measured in the range of $0.7 < p_{bound} < 1.0$, unless one can be certain of the exact population of free protein.

Impact of RDC Titration on the Accuracy of Molecular Interfaces in Weak Complexes.

Understanding the interaction structure of an organism's proteome is one of the key objectives of the many genomic and proteomic initiatives realized over recent decades. A complete understanding of molecular biology can only be derived from an atomic-level description of molecular interactions that control many cellular processes and are crucial for biological function. Weak molecular interactions are important for a vast range of cellular events, such as transcription and replication, signal transduction, transient formation of encounter complexes and assembly of protein complexes. NMR spectroscopy is the technique of choice for studying weak protein-protein, protein-nucleic acid or nucleic acid-nucleic acid interactions, characterizing transiently populated molecular complexes at atomic resolution. RDCs are powerful constraints that can be used to describe intermolecular orientation, but great care needs to exercised when using these constraints for the study of weakly interacting proteins. In this study we demonstrate that even in the case of weakly populated free forms of either protein, incorrect orientational information can be aliased into inaccurate position of the proteins in the subsequently determined complex, a scenario that is not detectable from the raw RDC data. The importance of this observation should not be underestimated for the understanding of molecular interaction. The example described here clearly demonstrates that under experimental conditions where only 80% of the protein is in the complex, the relative orientation of the entire protein can be incorrectly determined, incurring errors of up to 40-50°. These errors cannot be detected, giving back-calculated restraints that are in perfect agreement with the data, unless the titration procedure introduced here is applied. The consequences of this source of error are important, potentially leading to completely erroneous interpretation of the physical forces controlling molecular association and dissociation, for example electrostatic, hydrophobic or hydrogen bonding interactions, quite simply because the wrong amino acids will be aligned in the opposing interfaces. The fact that

interpretation of data from mixtures of free and bound proteins, and the resulting incorrect structures, cannot be detected from the raw data unless the RDC titration approach is applied, may explain why methods have not been developed to address this problem until now.

CONCLUSIONS

We have therefore developed a robust approach to the exploitation of the unique orientational information available from RDCs in the case of weak interaction and rapid exchange between free and bound forms. We demonstrate that the measurement of RDC changes upon titration of one partner into the equilibrium mixture leads to accurate determination of bound forms for both partners that are otherwise experimentally unattainable. We further develop analytical methods that guarantee the robustness of the approach by accurately adjusting the effective level of alignment of bound and free forms at all titration points. The method is applicable to a large number of proteins that can be studied by classical chemical shift perturbation, as long as both proteins and complex can be successfully aligned in the same medium, and that the complex can also be aligned. This approach provides complementary conformational restraints to those available from intermolecular contact restraints within a reasonable experimental period (3-4 days). We expect that this technique will extend the already remarkable and unique ability of NMR to determine the binding modes of weak complexes and expect this approach to contribute further to our understanding of diverse interactomes and thereby the molecular basis of cellular processes in different organisms.

ACKNOWLEDGMENTS

Grant BIO2005-04650 from the Spanish Ministry of Education and Science (MEC), the Commisariat à l'Energie Atomique, the French Centre National pour la Recherche Scientifique, the Université Joseph Fourier, Grenoble, the French Research Ministry through ANR NT05-4_42781 (M.B.), ANR JCJC05-0077 (B.B.). J.L.O.R. and N.A.J.v.N. are recipients of a FPU and Ramón y Cajal research contract from the MEC, respectively. A.I.A. is a recipient of a return grant of the Junta de Andalucia. M.R.J is supported by Lundbeckfonden and a long-term EMBO fellowship. We thank Varian for providing double-labelled ubiquitin. 600 MHz spectra recorded at Centre for Scientific Instrumentation (CIC) of the University of Granada and Institut de Biologie Structurale Grenoble. Financial support by the Access to Research Infrastructures activity in the 6th Framework Programme of the EC (Contract # RII3-026145, EU-NMR) for conducting the research at the RALF-NMR facility is gratefully acknowledged. We thank Adrien Favier,

Ewen Lescop and Rodolfo Rasia for pulse sequence development and spectrometer support.

Supporting Information Available. Methods, pulse sequence and structure calculation procedures. Tables showing AIRs and simulated alignment tensors. Figures of initial conformers, reproduction of experimental data compared to known structures of SH3 and ubiquitin.

REFERENCES

1. Chandonia, JM and Brenner, SE (2006) The impact of structural genomics: expectations and outcomes. *Science* **311**, 347–351.
2. Levitt, M Growth of novel protein structural data. (2007) *Proc. Natl. Acad. Sci. USA* **104**, 3183–3188.
3. Kelly, WP and Stumpf, MPH (2008) Protein–protein interactions, from global to local analyses *Current Opinion in Biotechnology* **19**,396–403.
4. Devos, D and Russell, RB (2007) A more complete, complexed and structured interactome. *Curr. Opin. Struct. Biol.* **17**, 370-377.
5. Levy, ED and Pereira-Leal, JB (2008) Evolution and dynamics of protein interactions and networks. *Curr. Opin. Struct. Biol.* **18**, 349-357.
6. Bader, S, Kühner, S and Gavin, A-C. (2008) Interaction networks for systems biology. *FEBS Lett* **582**, 1220-1224.
7. Nooren, I.M.A. and Thornton, J.M. (2003) Diversity of protein-protein interactions. *EMBO J.* **22** , 3486-3492.
8. Vaynberg, J and Qin, J (2006) Weak protein-protein interactions as probed by NMR spectroscopy. *Trends in Biotechnology* **24**, 22-27.
9. Tang, C, Iwahara, J and Clore GM (2006) Visualization of transient encounter complexes in protein-protein association. *Nature* **444**, 383-386.
10. Takeuchi, K. and Wagner, G. (2006) NMR Studies of Protein Interactions. *Curr. Opin. Struct. Biol.* **16**,109–117.
11. Tjandra, N and Bax, (1997) A Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, **278**, 1697-1697
12. Prestegard, JH, Bougault, CM and Kishore, AI. (2004) Residual dipolar couplings in structure determination of biomolecules. *Chem Rev.* **104**, 3519-3540.
13. Garrett, DS, Seok, Y-J, Peterkofsky, A, Gronenborn, AM and Clore GM (1999) Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nat. Struct. Biol.* **6** , 166–173.
14. Olejniczak, E.T. , Meadows, R.P., Wang, H., Cai, M., Nettesheim, D.G. and Fesik, S.W. (1999) Improved NMR Structures of

Protein/Ligand Complexes Using Residual Dipolar Couplings *J. Am. Chem. Soc.* **121**, 9249-9257.

15. Clore, GM (2000) Accurate and rapid docking of protein-protein complexes on the basis of intermolecular nuclear overhauser enhancement data and dipolar couplings by rigid body minimization. *Proc. Natl Acad. Sci.* **97**, 9021-9026.
16. McCoy, MA and Wyss, DF (2002) Structures of protein-protein complexes are docked using only NMR restraints from residual dipolar coupling and chemical shift perturbations. *J. Am. Chem. Soc*. **124**, 2104–2105.
17. Clore, GM and Schwieters, CD (2003) Docking of protein-protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from 1H/15N chemical shift mapping and backbone 15N-1H residual dipolar couplings using conjoined rigid body/torsion angle dynamics. *J. Am. Chem. Soc*. **125**, 2902–2912.
18. Bolon, PJ, Al-Hashimi, HM and Prestegard, JH (1999) Residual dipolar coupling derived orientational constraints on ligand geometry in a 53 kDa protein-ligand complex. *J. Mol. Biol.* **293**, 107-115.
19. Koenig, BW *et al* Structure and orientation of a G protein fragment in the receptor bound state from residual dipolar couplings. *J. Mol. Biol.* **322**, 441-461.
20. Jain, N.U.; Noble, S. and Prestegard, J.H. (2002) Structural characterization of a mannose-binding protein–trimannoside complex using residual dipolar couplings *J. Mol. Biol.* **326**, 451-462 (2003)
21. Jain, N.U., Wyckoff, T.J.O., Raetz, C.R.H. and Prestegard, J.H. (2004) Rapid Analysis of Large Protein–Protein Complexes Using NMR-derived Orientational Constraints: The 95kDa Complex of LpxA with Acyl Carrier Protein. *J. Mol. Biol.* **343**, 1379-1389.
22. Golovanov AP, Blankley RT, Avis JM and Bermel W. (2007) Isotopically discriminated NMR spectroscopy: a tool for investigating complex protein interactions in vitro. *J Am Chem Soc* **129**, 6528-6535.
23. Ciechanover, A. The ubiquitin-proteasome pathway: on protein death and cell life. (1998) *EMBO J.* **17**, 7151-7160.
24. Hicke L, Schubert, HL and Hill, CP Ubiquitin-binding domains. (2005) *Nat. Rev. Mol. Cell Biol.* **6**, 610.
25. Stamenova, S.D., French, M.E., He, Y., Francis, S.A., Kramer, Z.B. and Hicke, L. Ubiquitin binds to and regulates a subset of SH3 domains. (2007) *Mol. Cell* **25**, 273-284.
26. He, Y, Hicke, L and Radhakrishnan, I Structural basis for ubiquitin recognition by SH3 domains. (2007) *J. Mol. Biol.* **273**, 190-196.
27. Ortega Roldan, J.L., Ab, E., Romero Romero, M.L., Ora, A., Lopez Mayorga, O., Azuaga, A.I. and van Nuland, N.A.J. (2007) The

high resolution NMR structure of the third SH3 domain of CD2AP. *J. Biomol. NMR* **39**, 331-336
28. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes *J. Biomol. NMR* **6**, 277-293
29. Johnson, B.A. and Blevins, R.A. (1994) NMRView - a computer-program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603-614
30. Ruckert, M. and Otting, G. (2000) Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments *J.Am.Chem.Soc.* **122**, 7793-7797.
31. Schanda, P., Van Melckebeke, H. and Brutscher, B. (2006) Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. *J. Am. Chem. Soc.* **128**, 9042-9043
32. Lescop, E, Schanda, P and Brutscher, B. (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson*. **187**, 163-169.
33. Goddard, T.D., Kneller, D.G. **2003** *University of California*
34. Brutscher, B (2001) Accurate measurement of small spin-spin couplings in partially aligned molecules using a novel J-mismatch compensated spin-state-selection filter. *J. Magn. Reson*. **151**, 332-338.
35. Hus, JC, Marion, D and Blackledge, M. (2000) De novo determination of protein structure by NMR using orientational and long-range order restraints. *J.Mol.Biol.* **298**, 927-936.
36. 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. and Warren GL. (1998) Crystallography and NMR system (CNS): A new software system for macromolecular structure determination. *Acta Cryst* **D54**, 905–921
37. Dominguez, C, Boelens, R and Bonvin, AMJJ (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731-1737.
38. Sibille, N., Pardi, A., Simorre, J.P. and Blackledge, M. (2001) Refinement of Local and Long-Range Structural Order in Theophylline-Binding RNA using ^{13}C-^{1}H Residual Dipolar Couplings and Restrained Molecular Dynamics. *J.Am.Chem.Soc.* 123, 12135-12146
39. Sibille, N., Blackledge, M., Brutscher, B., Covès, J. and Bersch, B (2005) Solution structure of the sulfite reductase flavodoxin-like domain from Escherichia coli *Biochemistry* 44, 9086-9095
40. Zuiderweg, ER (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1-7.
41. Bonvin, AMJJ (2006) Flexible protein-protein docking. *Curr. Opin. Struct. Biol.* **16**, 194-200.

42. Van Dijk, ADJ and Bonvin, AMJJ (2006) Solvated docking, introducing water into the modelling of biomolecular complexes. *Bioinformatics* **22**, 2340-2347.
43. Cornilescu G, Marquardt JL, Ottiger M and Bax A. (1998) Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase *J. Am. Chem. Soc*., **120**, 6836-6837.
44. Dosset, P, Hus, J-C, Marion, D and Blackledge, M. (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings *J. Biomol. NMR* **20**, 223-231.
45. Zweckstetter, M and Bax, A. (2002) Evaluation of uncertainty in alignment tensors obtained from dipolar couplings *J. Biomol. NMR* **23**, 127-1377.
46. Bernado, P and Blackledge, M. (2004) Anisotropic small amplitude Peptide plane dynamics in proteins from residual dipolar couplings. *J. Am. Chem. Soc.* **126**, 4907-4920.
47. Bezsonova, I., Bruce, M.C., Wiesner, S., Lin, H., Rotin, D. and Forman-Kay, J.D. (2008) Interactions between the three CIN85 SH3 domains and ubiquitin: implications for CIN85 ubiquitination. *Biochemistry* **47**, 8937-8949.
48. Korzhnev DM, Bezsonova I, Lee S, Chalikian TV and Kay LE. (2009) Alternate Binding Modes for a Ubiquitin-SH3 Domain Interaction Studied by NMR Spectroscopy. *J.Mol.Biol. In Press.*

SUPPORTING INFORMATION

Methods

Protein expression and purification

For clarity, we use the residue numbering of CD2AP SH3-C of PDB entry 2JTE throughout the text. Unlabelled, 15N-labelled and ^{15}N,^{13}C-labelled CD2AP SH3-C was obtained as described[1]. Unlabelled and ^{15}N-labelled ubiquitin was purchased from both Cortecnet and Spectra Stable Isotopes. $^{15}N/^{13}C$-labelled ubiquitin was kindly provided by Varian Inc.

Protein concentrations were determined by absorption measurements at 280 nm using an extinction coefficient of 13980 and 1450 $cm^{-1}\cdot M^{-1}$ for SH3-C and ubiquitin, respectively, determined using the ProtParam algorithm (www.expasy.ch).

NMR chemical shift perturbation

All NMR titration experiments were performed at 25oC on a Varian NMR Direct- Drive Systems 600 MHz spectrometer (^1H frequency of 600.25 MHz) equipped with a triple-resonance PFG-XYZ probe. CD2AP SH3-C and ubiquitin samples were prepared for NMR experiments in 93% H2O/7% D2O, 50 mM NaPi, 1 mM DTT at pH 6.0. The backbone amide and ^{15}N frequencies of CD2AP SH3-C under the above conditions, previously assigned at pH 2.0.[1] were obtained first by comparing 2D ^1H-^{15}N HSQC spectra at pH 2.0, 3.0, 6.0 and 7.0 and confirmed by a single HNCACB triple resonance experiment acquired on $^{15}N/^{13}C$-labelled CD2AP SH3-C at pH 6.0. A HNCACB triple resonance spectrum was also recorded on a $^{13}C/^{15}N$-labelled ubiquitin to confirm backbone assignment at pH 6.0.

The SH3-binding site on ubiquitin was obtained by titrating with increasing amounts of unlabeled CD2AP SH3-C domain into a 0.25 mM ^{15}N-ubiquitin sample at pH 6.0, 25 °C. Similarly, the ubiquitin-binding site on CD2AP SH3-C was obtained by titrating with increasing amounts of unlabeled ubiquitin into a 0.25 mM ^{15}N-SH3-C sample under the same conditions. The progress of the titrations was monitored by recording one-dimensional ^1H and two-dimensional ^1H-^{15}N HSQC spectra.

The magnitude of the chemical shift deviations (Δδ) were calculated using the equation:

$$\Delta\delta = \sqrt{\left(\Delta\delta_{HN}\right)^2 + \left(\frac{\Delta\delta_N}{6.51}\right)^2} \qquad (6)$$

where the difference in chemical shift is that between the bound and free

forms of the different proteins. All NMR data were processed using NMRPipe[2] and analyzed by NMRView.[3]

NMR Spectroscopy

Assignment of SH3-C in the complex was confirmed by a single HNCACB triple resonance experiment acquired on ^{15}N,^{13}C-labelled CD2AP SH3-C in complex with unlabelled ubiquitin (molar ratio SH3-C:Ubiquitin of 1:3) at pH 6.0. In order to obtain atom-specific ambiguous interaction restraints (AIRs, see below), we recorded the following experiments on both free SH3-C as well as in complex with ubiquitin: HNCO, HNCACB, HBHA(CO)NH and aromatic (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE. These experiments thus provided us with chemical shift perturbations upon ubiquitin binding for all ^{1}H, ^{15}N, $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{1}H_{\beta}$, $^{13}C_{\beta}$, $^{13}C'$ atoms and the $^{1}H_{\delta}$ and $^{1}H_{\varepsilon}$ atoms of aromatic and Asn and Gln residues.

Measurement of Residual Dipolar Couplings

RDCs were measured in samples partially aligned in a liquid-crystalline medium consisting of a mixture of 5% penta-ethyleneglycol monododecyl ether ($C_{12}E_5$) and hexanol.[4] For the $^{13}C/^{15}N$-labelled protein (SH3) in the free form or in diverse mixtures of free and complex, a set of 4 different RDCs ($^{1}D_{NH}$, $^{1}D_{CaC'}$, $^{2}D_{HNC'}$ and $^{1}D_{CaHa}$) was measured per sample using 3D BEST-type HNCO or HNCOCA experiments[5, 6]. Coupling constants were obtained from line splittings in the ^{13}C dimension using the nmrPipe nlinLS fitting routine or using Sparky[7].For the ^{15}N- labelled protein (ubiquitin) in free or in the complex, $^{1}D_{NH}$ were measured from a pair of spin-state-selected ^{1}H-^{15}N correlation spectra recorded using the pulse sequence shown in figure S4. The pulse sequence uses a DIPSAP filter[8] for J-mismatch compensated spin-state selection, and the BEST concept for longitudinal relaxation and sensitivity enhancement. In addition, signals from the doubly ($^{13}C/^{15}N$) labeled binding partner are removed by additional transfer steps from ^{15}N to ^{13}CO. After this transfer step the presence of orthogonal coherences for the spin systems from ^{15}N- only and $^{13}C/^{15}N$ labeled proteins is exploited to suppress the unwanted signals by means of pulses field gradients and phase cycling. Total measurement time for one titration point is approximately 1 day. Sweep widths and data points used in different experiments:

	1H sw	13Csw	15Nsw	complex points(1Hx13Cx15N)	nt
3D HNCO_jnh:	7530.120	1170	2199.978	1024x120x120	2
3D HNCOCA_jcaha	7530.120	3000	2199.978	1024x120x120	4
3D HNCO_jcoh	7530.120	1700	2199.978	1024x120x100	4
3D HNCO_jcc	7530.120	1700	2199.978	1024x320x100	2
13CfiltDIPSAP	7530.120	------	2200	602x600	16

Structural model of the SH3-C domain of CD2AP in complex with ubiquitin using HADDOCK

Docking of the CD2AP SH3-C/Ubiquitin complex was performed using the software HADDOCK2.0[9] in combination with CNS[10] based on the chemical shift perturbation data obtained in this work. The starting structures for the docking were the 10 lowest- energy RDC-refined NMR structures of CD2AP SH3-C obtained in this work and the X-ray structure of ubiquitin (PDB entry 1UBQ)[11]. Active and passive residues defined for HADDOCK were chosen based on the chemical shift perturbation data and solvent accessibility. We first selected all the residues in ubiquitin having a combined backbone ^{1}H-^{15}N chemical perturbation upon complex formation higher than the mean chemical shift perturbation (0.05 ppm). We then selected ^{1}H, ^{15}N, $^{1}H_{\alpha}$, $^{13}C_{\alpha}$, $^{1}H_{\beta}$, $^{13}C_{\beta}$, $^{13}C'$ atoms and the $^{1}H_{\delta}$ and $^{1}H_{\varepsilon}$ atoms of aromatic and Asn and Gln residues in SH3-C showing chemical shift differences between the free and bound form bigger than the mean plus one standard deviation. We calculated the solvent accessibility using the program NACCESS[12], which is based upon the Lee & Richards algorithm[13], over the ensemble of SH3-C structures and the single ubiquitin X-Ray structure and selected as active residues all the amino acids showing an average relative solvent accessibility for the backbone and/or side-chains higher than 40%. We then selected all surface neighbors amino acids having a high solvent accessibility (>40%) as passive residues. These selection criteria resulted in active residues 6, 8, 42, 44, 46, 47, 48, 68, 71, 72, 73, 75 and 76 and passive residue 70 and 74 in ubiquitin. A 2 Å distance was used to define the ambiguous interaction restraints (AIR) for ubiquitin. Atom-specific AIRs for SH3-C involving active residues 15, 18,19, 21, 36-38, 40, 42, 43 and 56-59 are collected in Table S1 using a 6 Å distance for all protons and a 7 Å distance for ^{15}N and ^{13}C atoms. Residues 20 and 39 were included as passive residues.

Semi-flexible regions for the two proteins were defined as residues 12-24, 35-43 and 50-59 for SH3-C and residues 6-10, 40-50 and 68-76 for ubiquitin. During the rigid body energy minimization, 1000 structures were calculated. The 200 best solutions based on the intermolecular energy structure were used for the semi-flexible simulated annealing followed by refinement in explicit water. Finally, the solutions were clustered using a 3 Å rmsd after superposition on the backbone of the residues in the above defined semi-flexible regions. From each of the 10 resulting clusters, one

structure was then used as input for the combined chemical shift perturbation - RDC protocol.

RDC Refinement of the SH3-C domain of CD2AP in complex with ubiquitin

The program SCULPTOR[14] has recently been developed as an addition to the program CNS. The refinement protocol involves a restrained MD calculation using the standard CNS force field. Starting structures were taken from a selection of 10 structures determined using HADDOCK[9] based on ambiguous intermolecular restraints (AIRs). Initial sampling was further increased by including a sampling period of 10ps at 700K that allows for the SH3 domain to reorient freely without RDCs or AIR restraints. Following this both ubiquitin and SH3 conformations are fixed and the alignment tensor is allowed to evolve freely to determine initial estimates of the alignment tensor on the SH3 structure (there are four RDC types available for this structure). Both molecules and tensors are then freed with the initial sampling period at 1000K for 5ps during which time both AIRS and RDCs are scaled from 0.1% to 100% of their final values. The backbone conformation of the segment (1-70) of ubiquitin is restrained to its initial coordinates using a harmonic potential, while the experimentally measured NOes from the free form of SH3 are used as restraints. AIR restraints are used as described in the supporting information. Sampling of restraints is followed by a 5ps sampling stage and slow cooling over 5ps to 100K and energy minimization. The protocol is repeated 30 times for each starting structure.

Supplementary tables

Table S1: Atom-specific ambiguous interaction restraints for SH3-C in complex with ubiquitin.

Residue in SH3	^{15}N	^{1}HN	^{13}Cα	^{1}Hα	^{13}CO	^{13}Cβ	^{1}Hβ	^{1}Hδ	^{1}Hε
Tyr15	X		X	X	X			X	X
Thr18		X	X						
Asn19				X		X	X	X	
Asp21			X			X	X		
Lys36				X	X				
Glu37	X					X			
Thr38		X			X				
Glu40			X			X			
Gly42					X				
Trp43				X		X			
Pro56			X				X		
Asp57				X			X		
Asn58								X	
Phe59				X		X	X	X	X

Table S2. Alignment tensors of different mixtures between two different simulated tensors.

p	0.0[a]	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
A_a (10^{-4})	12.8	11.8	11.0	10.3	9.8	-9.9	-10.8	-11.9	-13.2	-14.6	-16.0
A_r (10^{-4})	4.0	3.9	4.1	4.7	5.6	-6.1	-5.4	-4.7	-4.2	-4.1	-3.8
α (°)	107	113	120	126	129	52	52	52	55	59	63
β (°)	72	69	66	64	61	56	57	58	59	61	62
γ (°)	143	145	148	151	155	-88	-82	-78	-75	-73	-79

[a]*RDC data from the two limiting tensors (p=0.0 and 1.0) were simulated and co-added to mimic the case of free and bound form mixtures. The data were then fitted to the same structure, and the effective alignment tensors are shown here. The fit is perfect in each case.*

Table S3. Structural statistics for the 10 models of the CD2AP SH3-C:Ubiquitin complex[1]

Restraints statistics	
NOe violations > 0.5 Å	26 ± 9
NOe violations > 0.3 Å	42 ± 14
NOe violations > 0.1 Å	66 ± 23
AIR violations > 0.2 Å	0.3 ± 0.3
φ/ψ restraint violations > 5°	2.0 ± 0.20
Reduced Chi2 (χ^2/N) from experimental residual dipolar coupling restraints:[2]	
SH3	
C^{α}-H^{α}	(42) 1.5 ± 0.2
N-H^N	(56) 0.8 ± 0.1
C'-H^N	(55) 1.1 ± 0.1
Cα-C'	(53) 1.2 ± 0.1
Ubiquitin	
N-H^N	(56) 0.9 ± 0.1
Alignment tensor eigenvalues:	
A_a	12.9 ± 0.1 10^{-4}
A_r	3.7 ± 0.1 10^{-4}
Deviation from the idealized covalent geometry	
Bonds [Å]	0.0025 ± 0.0001
Angles [°]	0.43 ± 0.01
Impropers [°]	0.49 ± 0.02
CNS Interaction Energies (Kcal/mol)	
E_{elec}	-445 ± 104 kcal.mol^{-1}
E_{vdw}	-756 ± 18 kcal.mol^{-1}
Coordinate precision (Å)[3]	
Backbone N, Cα, C'	1.0 ± 0.1
Ramachandran plot[4]	
Most favored regions (%)	87.0
Additional allowed regions (%)	10.4
Generously allowed regions (%)	2.6
Disallowed regions (%)	0.0

[1]The statistics is obtained from an ensemble of ten models for the SH3-C:Ubiquitin complex with the lowest target function. [2]Restraint statistics reported for the refined complex. [3]Coordinate precision is given as the pair-wise Cartesian coordinate Root Mean Square Deviations from the average structure of the entire complex over the ensemble (SH3 and Ubiquitin). [4]Values obtained from the PROCHECK analysis over all residues.

Supplementary figures

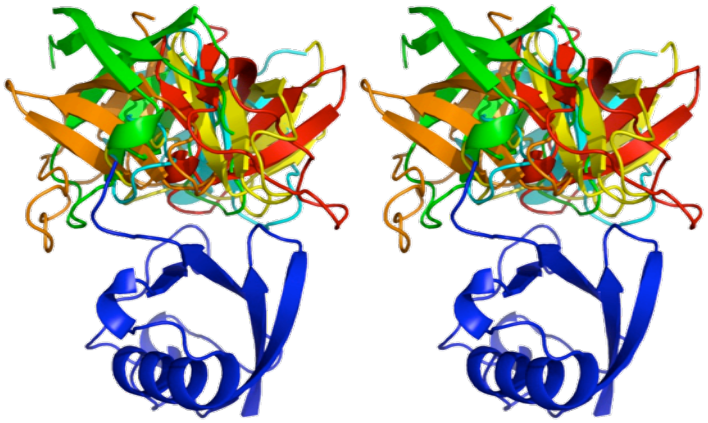


Figure S1. Representative cartoon presentation of starting structures from HADDOCK. *Stereo representation of 5 out of the 10 HADDOCK structures that were used as starting structures in the combined CS-RDC protocol. Structures were superimposed on the backbone atoms of residues 4-73 of the ubiquitin molecule. For clarity only one ubiquitin molecule is presented in blue, while the 5 CD2AP SH3-C molecules are in different colours.*

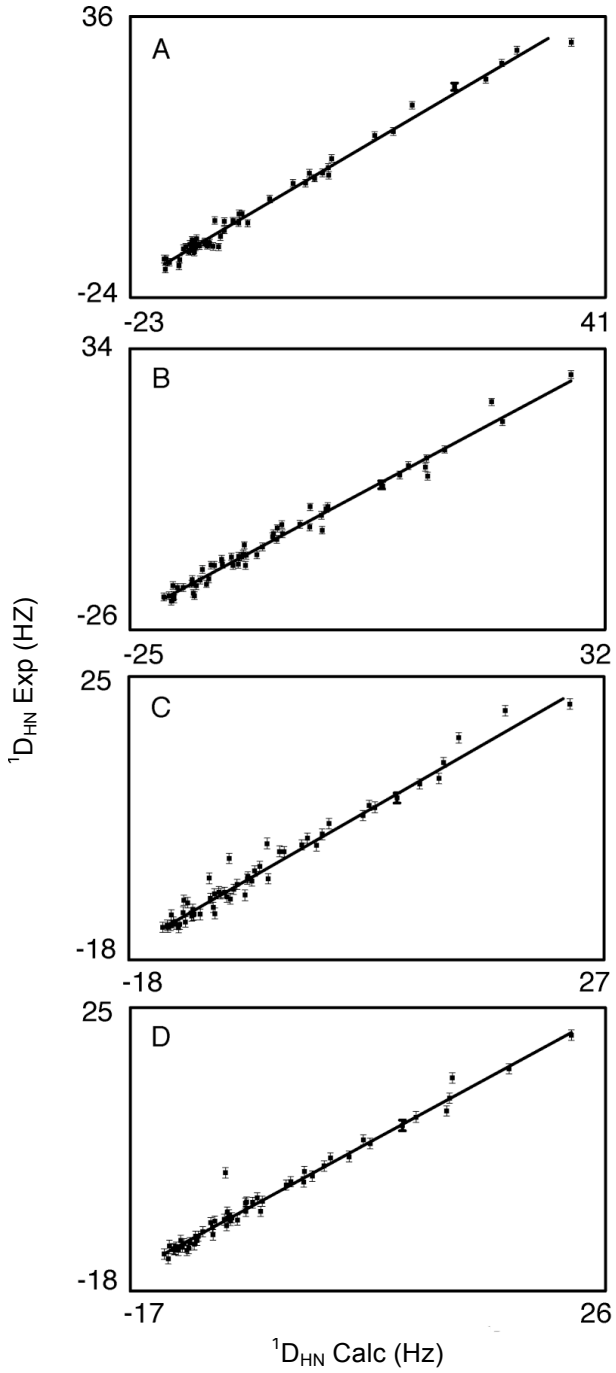


Figure S2. Fits of experimental RDC data sets to the 1d3z conformation of ubiquitin. *N-HN RDCs are reproduced by the structure for all four data sets using the best fitting alignment tensors (table 1). (A) free form, (B) mixture m1, (C) mixture m2, (D) mixture m3.*

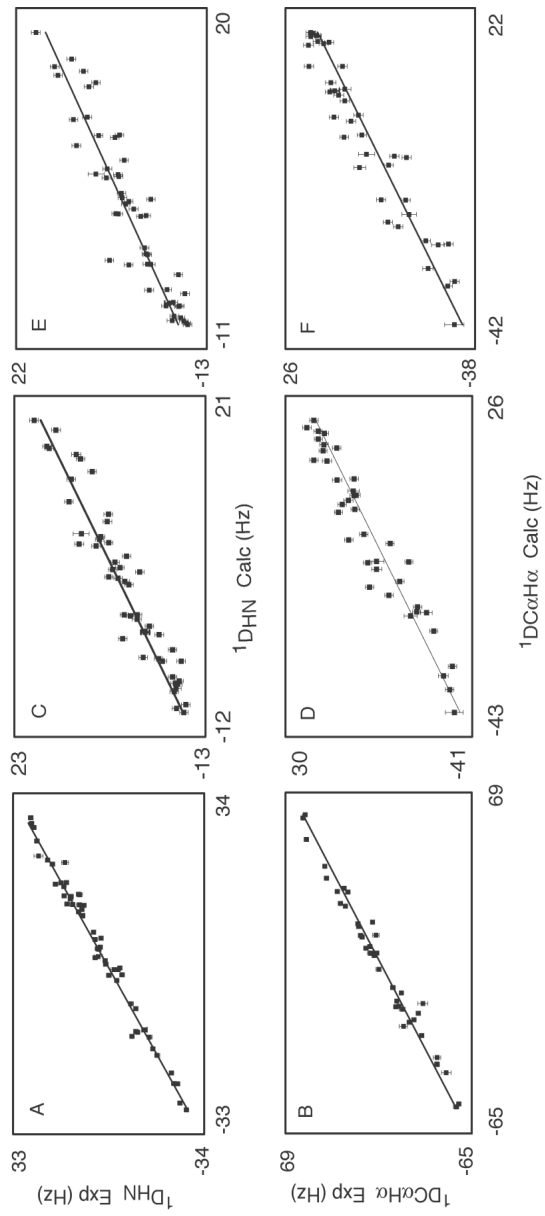


Figure S3. Fits of experimental RDC data sets to the RDC refined free SH3 conformation. (A, C, E) *Comparison of N-HN RDCs to mixtures m1, m2 and m3 respectively for the best fitting alignment tensors (table 1) .* ***(B, D, F)*** *Comparison of CαHα RDCs to mixtures m1, m2 and m3 respectively for the best fitting alignment tensors (table 1). While inferior agreement of experimental data with the SH3 free- form structure may indicate that more structural rearrangements are taking place in this protein upon formation of the complex, it may also reflect the lower effective resolution of the free-form structure than is available for ubiquitin.*

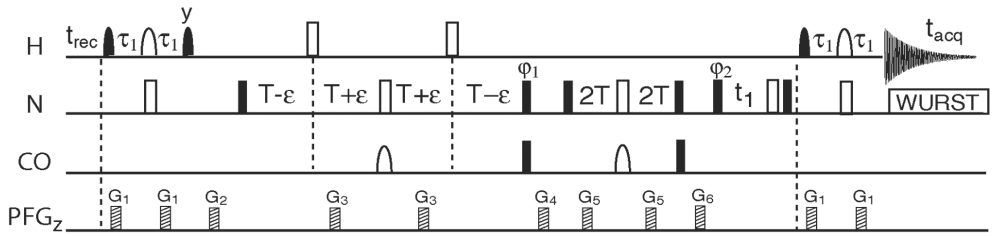


Figure S4: BEST-type pulse sequence for the measurement of (1JNH+1DNH) coupling constants in a mixture of 15N- and 13C/15N-labelled proteins. *Only signals from the 15N-labelled protein are detected in this experiment, while all signals from the doubly labeled protein are suppressed by a ^{15}N-^{13}CO transfer filter. Filled and open pulse symbols indicate 90° and 180° rf pulses. Unless indicated, all pulses are applied with phase x. All selective 1H pulses are centered at 8.2 ppm, covering a bandwidth of 4.0 ppm, with the following shapes: EBURP26 for excitation, REBURP6 for refocusing, and time-reversed EBURP26 for flip back purposes. The additional 180° 1H pulses applied during the DIPSAP filter (open squares) have the shape of BIP-720-50-20 pulses7. CO pulses have the shape of the center lobe of a sinx/x function. The transfer delays are set to: τ1=1/4JHN ≈ 2.4 ms-δ/2, and T= 1/8JNCO ≈ 8.5 ms, with δ the length of the REBURP 180° pulse. Pulsed field gradients, G1–G8 are applied along the z-axis (PFGz) with durations of 200 μs to 2 ms and field strengths ranging from 5–40 G/cm. A 2-step phase cycle is applied by simultaneously changing the sign of φ1 and the receiver phase φrec. Quadrature detection in t1 is obtained by time-proportional phase incrementation of φ2 according to TPPI-States. For DIPSAP spin-state selection8 3 repetitions of the experiment are performed with the following settings: (A) ε=0, φ1 = x, (B) ε =1/8JHN ≈ 1.3 ms, φ1 = y, and (C) ε =1/4JHN ≈ 2.6 ms, φ1 = x. The spectra containing the upfield and downfield components of the 15N doublets are then calculated from the 3 recorded data sets: 0.73 A – 0.27 C ± B.*

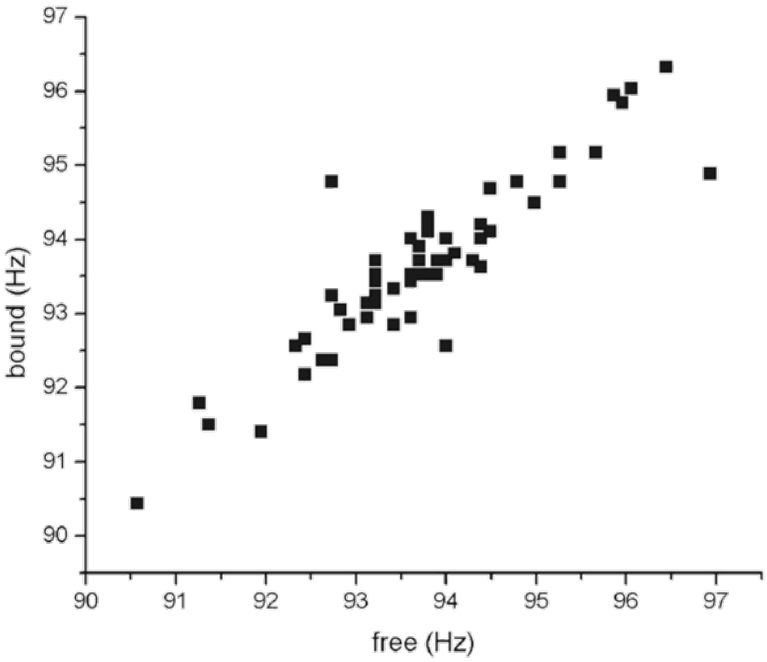


Figure S5. Correlation plot of 1H-15N scalar couplings in free and complexed form. *Correlation between scalar couplings measured in free SH3 versus scalar couplings measured in mixture m1. To create two of the mixtures (m2 and m3), ubiquitin was added to the aligned sample, so that scalar couplings were not re-measured. For mixture m1 scalar couplings were measured before adding the alignment medium. The high correlation between measured scalar couplings in the free and free/complex mixture m1 justifies the approximation used here.*

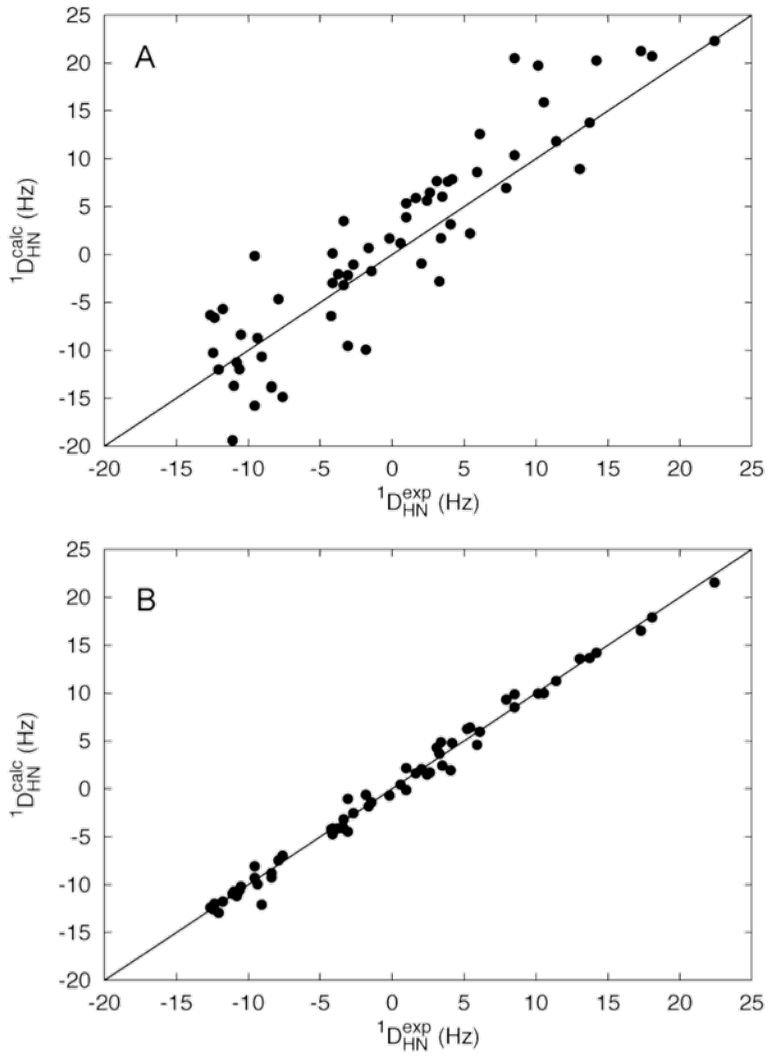


Figure S6. *The ability of the RDC titration procedure to determine 'free' data sets, compared to a simple extrapolation from a single mixture with the relative scaling determined from the D2O splitting measured in the free and bound samples. All 15N- 1HN RDCs were removed from mixture m3 and :*
(A) predicted on the basis of values measured at m0 and m1 using a standard extrapolation $\left(D_{ij}^{m_3} = \left(\lambda_1 p_3 D_{ij}^{m_1} - \lambda_{0,SH3}\left(p_3 - p_1\right)D_{ij}^{free}\right)\Big/\lambda_3 p_1\right)$. *The populations p are assumed to be accurately known in both cases and the scaling terms λ represent the measured D2O splitting (table 2).*
(B) the RDC titration analysis described in the article was repeated and experimental and 15N-1HN predicted values from mixture m3 were compared.

References

1. Ortega Roldan, J.L.; Ab, E.; Romero Romero, M.L.; Ora, A.; Lopez Mayorga, O.; Azuaga, A.I.; van Nuland, N.A.J. *J. Biomol. NMR* **2007** 39, 331-336
2. Delaglio, F.; Grzesiek, S.; Vuister, G.W.; Zhu, G.; Pfeifer, J.; Bax, A *J. Biomol. NMR* **1995** 6, 277-293
3. Johnson, B.A.; Blevins, R.A. *J. Biomol. NMR* **1994** 4, 603-614
4. Ruckert, M.; Otting, G. *J.Am.Chem.Soc.* **2000** 122, 7793-7797.
5. Schanda, P.; Van Melckebeke, H.; Brutscher, B. *J. Am. Chem. Soc.* **2006** 128, 9042-9043
6. Lescop, E.; Schanda, P.; Brutscher, B. *J. Magn. Reson*. **2007** 187, 163-169.
7. Goddard, T.D.; Kneller, D.G. **2003** *University of California*
8. Brutscher, B. *J. Magn. Reson*. **2001** 151, 332-338.
9. Dominguez C, Boelens R, Bonvin AM (2003) *J. Am. Chem. Soc.* 125: 1731-1737
10. 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 Cryst* D54: 905–921
11. Vijay-Kumar S, Bugg CE, Cook WJ Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol.* 194, 531–544 (1987)
12. Hubbard SJ, Thornton JM NACCESS. Department of Biochemistry and Molecular Biology, University College, London (1993)
13. Lee B, Richards FM The interpretation of protein structures: estimation of static accessibility *J. Mol. Biol.* 55, 379-400 (1971)
14. Sibille, N.; Pardi, A.; Simorre, J.P.; Blackledge, M. *J.Am.Chem.Soc.* **2001** 123, 12135-12146

Chapter 5

Distinct ubiquitin binding modes exhibited by SH3 domains of CD2AP: Molecular determinants and functional implications.

Submitted to the Journal of Biological Chemistry (2010) by Jose L. Ortega-Roldan, Salvador Casares, Malene Ringkjøbing-Jensen, Martin Blackledge, Ana I. Azuaga and Nico A.J. van Nuland

ABSTRACT

Ubiquitin is known to regulate a wide variety of cellular activities ranging from transcriptional regulation to cell signalling and membrane trafficking. Many cellular activities are mediated by mono- rather than poly-ubiquitin, and its functions are deciphered by various ubiquitin-binding-proteins. SH3 domains were found to constitute a new type of ubiquitin-binding domains. The structure of the complex between Sla1 SH3-3 and ubiquitin shows that the ubiquitin binding surface of the Sla1 SH3 domain overlaps largely with the canonical binding surface for proline-rich ligands. A key affinity and specificity determinant for ubiquitin-binding was appointed to Phe409 of Sla1, located at the heart of the hydrophobic interface in the complex. Moreover, a tyrosine residue at the equivalent position was found to abrogate ubiquitin binding. We recently showed that the third SH3 domain (SH3-C) of CD2AP binds ubiquitin in an alternative orientation. Here we show that the first and second SH3 domains of CD2AP bind ubiquitin in a similar orientation than Sla1 SH3-3 despite of their high sequence homology with the third domain of CD2AP. The structural work in combination with mutational analysis have allowed us to decipher the determinants of the SH3-C binding mode to ubiquitin. We also show that SH3 domains displaying this alternative binding mode interact with higher affinity to C-terminal extended versions of ubiquitin. We conclude that CD2AP SH3-C domain interaction with ubiquitin constitutes a new ubiquitin binding mode involved in a different cellular function, and thus changes the previously established mechanism of EGF-dependent CD2AP/CIN85 mono-ubiquitination.

INTRODUCTION

Adaptor molecules are non-catalytic polypeptides that contain one or more domains able to bind to other protein or non-protein ligands [1]. These molecules selectively control the spatial and temporal assembly of multi-protein complexes that transmit intracellular signals involved in regulation of cell growth, differentiation, migration and survival. The importance of such signalling networks is well understood for signal transduction induced by receptor tyrosine kinases (RTKs) [2].

Down regulation of RTKs is a critical step in modulating their activity [3]. This is a highly ordered and dynamic process controlled by the assembly of a large network of RTK-associated proteins as well as their post-translational modifications including phosphorylation and ubiquitination. The Cbl family of ubiquitin ligases plays a major role in receptor ubiquitination, receptor sorting for degradation and ultimately in cessation of receptor-induced signal transduction. Several lines of evidence support a role of CMS/CIN85 (Cas ligand with Multiple SH3 domains/Cbl-

Interacting protein of 85 kDa) in the regulation of Cbl-directed RTK down regulation [4-7]. CMS and its mouse homolog CD2AP (CD2 associated protein), shares an identical overall domain organisation, containing three SH3 (Src-Homology) domains, a proline-rich region and a coiled-coiled domain, and high sequence identity with CIN85, and hence, it has been assumed to belong to the same family of ubiquitously expressed adaptor molecules and elicit similar biological functions.

Among others, CD2, nephrine, podocine and c-Cbl have so far been identified as natural targets for CD2AP. Intermolecular interactions known until now are mostly mediated via the three N-terminal SH3 domains, named A, B and C. These three SH3 domains share higher similarity among themselves than to any other SH3 domains, suggesting that they may have overlapping specificities in binding [8].

SH3 domains form a highly conserved family of domains, but their amino acid composition varies at a few key sites, allowing for a wide range of molecular targets. Similar to SH3 domains, ubiquitin-binding domains (UBDs) are found in proteins with different biological function. Structures of many complexes between UBDs and ubiquitin have been determined either by X-ray or high-resolution NMR (reviewed in [9, 10]). Recently, SH3 domains were identified as a new class of UBDs [12]. Since then, several low- and high-resolution structural studies have raised a number of unanswered questions on the determinants of the specificity of the interaction between SH3 domains and ubiquitin, [11-14, 17, 18]. Moreover, mechanistic differences have been proposed for ubiquitin binding to SH3 domains involved in the immune signalling pathway [11].

We recently used NMR residual dipolar couplings (RDCs) to obtain the solution structure of the complex between ubiquitin and the third (SH3-C) domain of CD2AP [12]. In this work, we use RDCs together with chemical shift perturbations to obtain the solution structure of the complex between ubiquitin and the first (SH3-A) domain of CD2AP, and demonstrate that the two domains from the same adaptor protein show alternative modes of ubiquitin binding. We show that the CD2AP SH3-C domain binds to ubiquitin mainly involving polar residues positioned in the RT and n-Src loop and in the 3_{10} helix of the protein, and does not involve some of the conserved aromatic residues that are important for recognition of poly-proline sequences. This is in contrast to the CD2AP SH3-A domain, where the binding interface for ubiquitin is found to completely overlap with that for poly-proline sequences, similar to the Sla1 SH3-3 and the CIN85 SH3-C domains [13-15]. As a consequence, the mutation of phenylalanine 324 to tyrosine residue in CD2AP SH3-C does not abolish ubiquitin binding, opposite to the key affinity and specificity determinant appointed to this phenylalanine residue for ubiquitin binding by SH3 domains [13-15] and as

observed for CD2AP SH3-A, where the corresponding mutation of phenylalanine 53 to tyrosine abrogates ubiquitin binding.

The differences in binding mode of SH3 domains to ubiquitin might be related to different functions inside the cell. We suggest that the higher affinity observed for CD2AP SH3-C and other SH3 domains to ubiquitin that is rigidified by additional stretches of residues at its C-terminus, might be related to selective binding to ubiquitin molecules covalently bonded via its C-terminus in ubiquitinated proteins.

EXPERIMENTAL PROCEDURES

Protein expression and purification

Unlabelled, ^{15}N-labelled and ^{13}C ^{15}N CD2AP SH3-A and C were obtained as described in chapter 3. Unlabelled CD2AP SH3-B and CIN85 SH3-C were expressed as described in chapter 3. Unlabelled and ^{15}N-labelled His-tagged and non His-tagged Ubiquitin were purchased from Cortecnet or obtained as described in chapter 4. Protein concentrations were determined by absorption measurements at 280 nm using extinction coefficients of 9970, 12660, 13980, 13980 and 1450 M^{-1} cm^{-1} for CD2AP SH3-A, B and C, CIN85 SH3-C and Ubiquitin respectively, determined using the ProtParam algorithm (www.expasy.ch)

The plasmid encoding the mutant Phe324Tyr SH3-C protein was obtained by site-directed mutagenesis using QuickChange™ (Stratagene) kit in the plasmid pETM-11. The mutant protein was purified like the WT CD2AP SH3-C domain as described previously. Phe53Tyr CD2AP SH3-A mutant and CD2AP SH3-C Thr283Ala and Glu302Lys + Thr303 deletion mutants were purchased from Topgene and expressed similar to the WT domains. The DNA was sequenced and the molecular weight of the final purified protein was checked by mass spectrometry to confirm incorporation of every mutation.

NMR chemical shift perturbation.

All NMR titration experiments were recorded at 25°C on a Varian NMR Direct-Drive 600 MHz spectrometer (^{1}H frequency of 600.25 MHz) equipped with a triple-resonance PFG-XYZ probe. ^{15}N labelled-CD2AP SH3-A and C samples and ubiquitin were prepared in 92% H_2O/8% D_2O, 50 mM NaPi, 1 mM DTT at pH 6.0. A HNCACB triple resonance spectrum was recorded on a ^{13}C ^{15}N-labelled ubiquitin to confirm the backbone assignment at pH 6.0. CD2AP SH3-A and C backbone resonances were

previously assigned at pH 6.0 (BMRB accession numbers 16641 and 16643).

The ubiquitin-binding site on all SH3 domains was obtained by titrating with increasing amounts of unlabelled ubiquitin into a 0.2 mM ^{15}N-labelled SH3 sample. The SH3-binding site on ubiquitin was obtained by titrating with increasing amounts of unlabelled SH3 domain into a 0.25 mM ^{15}N-ubiquitin sample. The progress of the titrations was monitored by recording one-dimensional ^{1}H and two-dimensional ^{1}H-^{15}N HSQC spectra. The magnitude of the chemical shift deviations (Δδ) were calculated using the equation:

$$\Delta\delta = \sqrt{\left(\Delta\delta_{HN}\right)^2 + \left(\frac{\Delta\delta_N}{6.51}\right)^2}$$

where the difference in chemical shift is that between the bound and free forms of the different points. All NMR data were processed using NMRPipe [16] and analyzed by NMRView [17] and Sparky [18].

Measurement of RDCs and structure calculation

Partially aligned samples in a liquid-crystalline medium consisting of a 5% penta-ethyleneglycol monododecyl ether ($C_{12}E_5$)/hexanol mixture [19] were prepared for RDC measurement. RDCs were measured for the CD2AP SH3-C:Ubiquitin complex as described in [12]. D_{NH} couplings were measured in two partially aligned mixtures of CD2AP ^{15}N labelled SH3-A and unlabelled Ubiquitin and visa versa at 1:2.5 and 2.5:1 SH3-A:Ubiquitin ratios from a pair of spin-state-selected ^{1}H-^{15}N correlation spectra recorded using an DIPSAP [20] filter for J-mismatch compensated spin-state selection, and the BEST concept [21, 22] for longitudinal relaxation and sensitivity enhancement. Coupling constants were obtained from line splittings in the ^{15}N dimension using the NMRPipe nlinLS routine.

The structure calculation of the CD2AP SH3-A and C domains in complex with ubiquitin were carried out using the program SCULPTOR [23] as described in [12]. The ten lowest target function structures (combining RDC and AIR violation for the CD2AP SH3-A complex and RDC, AIR and NOe violation for the CD2AP SH3-C complex) were used for the final analysis. Validation of the structural quality was done using PROCHECK [24, 25]. H-bonds and non-bonded interactions were obtained from analysis done with the LIGPLOT software [26]. Salt-bridges were obtained from analysis using the EBI PISA web server [27].

Fluorescence binding experiments

Fluorescence measurements were made on a Varian Cary Eclipse spectrofluorimeter equipped with a temperature-controlled cell holder. Spectra were recorded at 25°C between 305 and 400 nm with an excitation wavelength of 298 nm to predominantly excite the tryptophan residues only present in the SH3 domain. Each spectrum was the result of 5 accumulations collected at a scan rate of 200 $nm \cdot min^{-1}$. First, the spectra were recorded on CD2AP SH3-C WT (30 μM) and F324Y (40 μM) and CD2AP SH3-A (20 μM) samples in the absence of ubiquitin. Spectral recording was repeated after the addition of a concentrated ubiquitin solution (760 μM) to an SH3:ubiquitin molar ratio of ~1:20. All spectra were corrected by subtracting a spectrum recorded on the solution without the SH3 domain, and then normalized by protein concentration.

Isothermal Titration Calorimetry (ITC)

ITC experiments were performed at 25°C in a high-precision MCS isothermal titration calorimeter and in a VP isothermal titration calorimeter (Microcal Inc., Northampton, USA). Prior to the experiments, WT and F324Y CD2AP SH3-C, as well as ubiquitin were extensively dialyzed at 4°C against the appropriate buffer (20 mM Cacodylate, pH 6.0). All buffers and solutions were filtered, degassed and equilibrated to the working temperature prior to each experiment. Typically, a 50 μM CD2AP SH3 solution in the calorimeter's cell was titrated with the ubiquitin solution (at a concentration of about 800 μM). Due to the relatively low binding affinities, titrations were made by using a profile of injection volumes of the ubiquitin solution in order to define the titration isotherm curve more accurately. As a blank, an independent experiment with only buffer in the calorimeter's cell was performed with the same ubiquitin solution to determine the corresponding heats of dilution. The dilution isotherm was subtracted from that obtained for the titration of the protein. The area under each peaks of the net thermogram was integrated to determine the heat produced by the binding event between the ubiquitin and the SH3 domain after each injection.

The binding constant, K_b, and the enthalpy change of binding, ΔH_b, were determined by fitting of the binding isotherms, using a model of binding to a single set of identical sites as described previously [28].

RESULTS

NMR titration experiments of CD2AP SH3-A and C domains with ubiquitin.

NMR chemical shifts of protein residues are highly sensitive to changes in the local environment, and chemical shift perturbations are therefore widely used to map intermolecular interfaces of protein complexes [29]. Titrations of ^{15}N-labeled ubiquitin with increasing amounts of SH3-A and C domains at 25°C and *visa versa* caused a selective shift of amide proton and nitrogen resonances of several ubiquitin and SH3-A (Figure 1A and B) and SH3-C residues (Figure 2A and B), indicating a specific union between the two proteins.

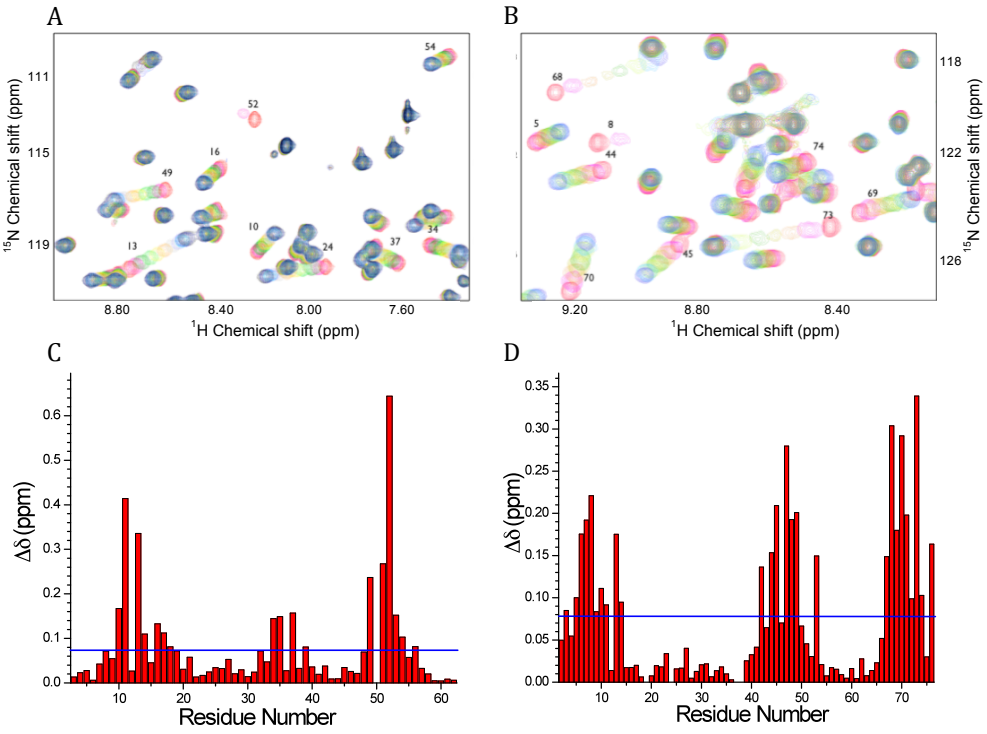


Figure 1: Monitoring the binding between ubiquitin and CD2AP SH3-A by NMR chemical shift perturbations. A. *^{1}H and ^{15}N Chemical shift changes upon titrating ubiquitin into a ^{15}N-labelled CD2AP SH3-A solution. Region of ^{1}H-^{15}N HSQC spectra recorded at increasing amounts of ubiquitin till a ratio of 1:1.6 (SH3:Ubi) (red to blue).* ***B.*** *^{1}H and ^{15}N Chemical shift changes upon titrating CD2AP SH3-A into a ^{15}N-labelled ubiquitin solution. Region of ^{1}H-^{15}N HSQC spectra recorded at increasing amounts of CD2AP SH3-A till a ratio of 1:1.4 (Ubi:SH3) (red to blue).* ***C and D.*** *Chemical shift deviations (Δδ) between the first and last titration point of the ubiquitin titration into ^{15}N-labelled CD2AP SH3-A and visa-versa, respectively. Blue solid lines indicate the mean chemical shift deviations that were used to define the active residues used in the structure calculation as AIR.*

The most significant changes in the chemical shift of the HSQC cross-peaks of the SH3-A domain are observed at the RT loop (residues 10-11, 13-14 and 16-18), the n-Src loop (residues 34-35), at the beginning of the β-III strand (residues 37 and 39) and at the 3_{10} helix and β-V strand (residues 49-56) (Figure 1C). While the titration profiles for SH3-A show intermediate exchange behaviour, fast exchange is observed in the case of SH3-C, indicating less tight binding for the third CD2AP SH3 domain.

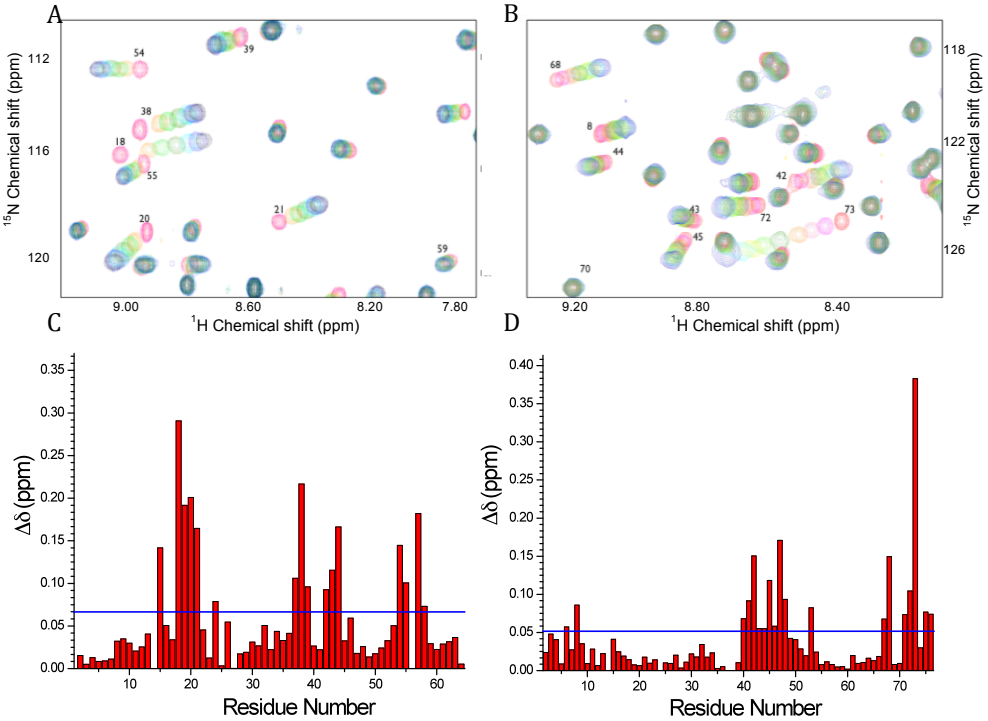


Figure 2: Monitoring the binding between ubiquitin and CD2AP SH3-C by NMR chemical shift perturbations. A. *^{1}H and ^{15}N Chemical shift changes upon titrating ubiquitin into a ^{15}N-labelled CD2AP SH3-C solution. Region of ^{1}H-^{15}N HSQC spectra recorded at increasing amounts of ubiquitin till a ratio of 1:2.4 (SH3:Ubi) (red to blue).* ***B.*** *^{1}H and ^{15}N Chemical shift changes upon titrating CD2AP SH3-C into a ^{15}N-labelled ubiquitin solution. Region of ^{1}H-^{15}N HSQC spectra recorded at increasing amounts of CD2AP SH3-C till a ratio of 1:2.5 (Ubi:SH3) (red to blue).* ***C and D.*** *Chemical shift deviations (Δδ) between the first and last titration point of the ubiquitin titration into ^{15}N-labelled CD2AP SH3-C and visa-versa, respectively. Blue solid lines indicate the mean chemical shift deviations that were used to define the active residues used in the structure calculation as AIR.*

For SH3-C substantial differences between the free and bound HSQC cross-peaks are found in the RT loop (residues 283-288), the n-Src loop (residues 302-304), the beginning of the βIII -strand (residues 307-309) and in residues 319-323 located at the beginning of the 3_{10} helix (Figure 2C).

Different chemical shift perturbation profiles are found when titrating the two domains into ^{15}N-labelled ubiquitin. Titration with unlabelled SH3-A causes considerable changes in the chemical shifts of residues in mainly three regions of the protein, around Leu8, around Ile44, a common binding region for most UBDs, and around Val70 (Figure 1D). These three regions are affected in a similar extent. In contrast, titration of ^{15}N ubiquitin with unlabelled SH3-C induces most significant changes in the Ile44 and Gly76 binding regions of ubiquitin, and very small changes are observed around Leu8 (Figure 2D).

In light of the different results obtained for the two SH3 domains of CD2AP, we decided to investigate if and how the second domain (SH3-B) would bind ubiquitin. Chemical shift perturbations observed when titration ^{15}N-labeled ubiquitin with CD2AP SH3-B are very similar to that of CD2AP SH3-A (Figure 3A), suggesting an analogous ubiquitin binding mode for the two N-terminal SH3 domains of CD2AP.

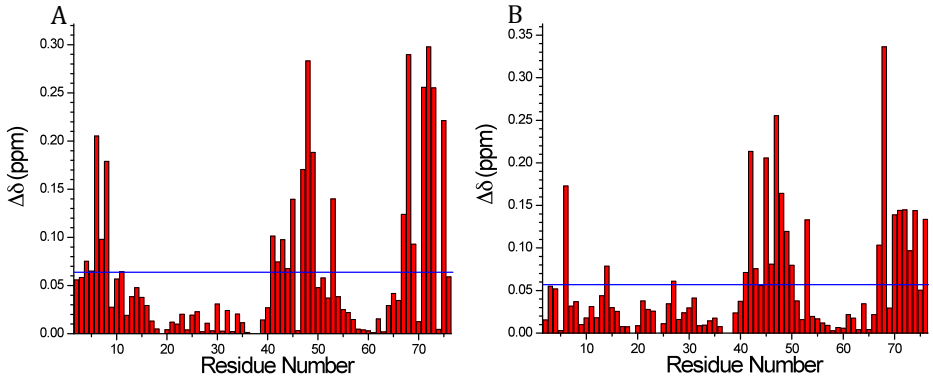


Figure 3: Chemical shift perturbations resulting from binding of ubiquitin to CD2AP SH3-B and CIN85 SH3-C domains. *Comparison of the chemical shift deviations (Δδ) between the first and last titration point of titrating CD2AP SH3-B (**A**) and CIN85 SH3-C (**B**) into ^{15}N-labelled ubiquitin corresponding to a ratio of 1:1.2 and 1:1.1 (SH3:Ubi) respectively. Blue solid lines indicate the mean chemical shift deviations.*

The three dimensional structure of the complex between the CD2AP SH3-A and C domains with Ubiquitin.

Despite of the high sequence homology between the three SH3 domains of CD2AP, the chemical shift perturbations seem to indicate structural differences between the different SH3-Ubiquitin complexes. We have determined the three dimensional solution structures of CD2AP SH3-A and SH3-C in complex with ubiquitin in order to understand the basis of these

differences and to compare with the previously determined structures of other SH3:Ubiquitin complexes (PDB entries 2JT4 and 2K6D). All backbone resonances of SH3-A and C and ubiquitin were assigned by following the shift of the previously assigned chemical shift of each cross-peak of the free forms upon titration and checked using 3D HNCACB, CBCA(CO)NH and HBHA(CO)NH experiments.

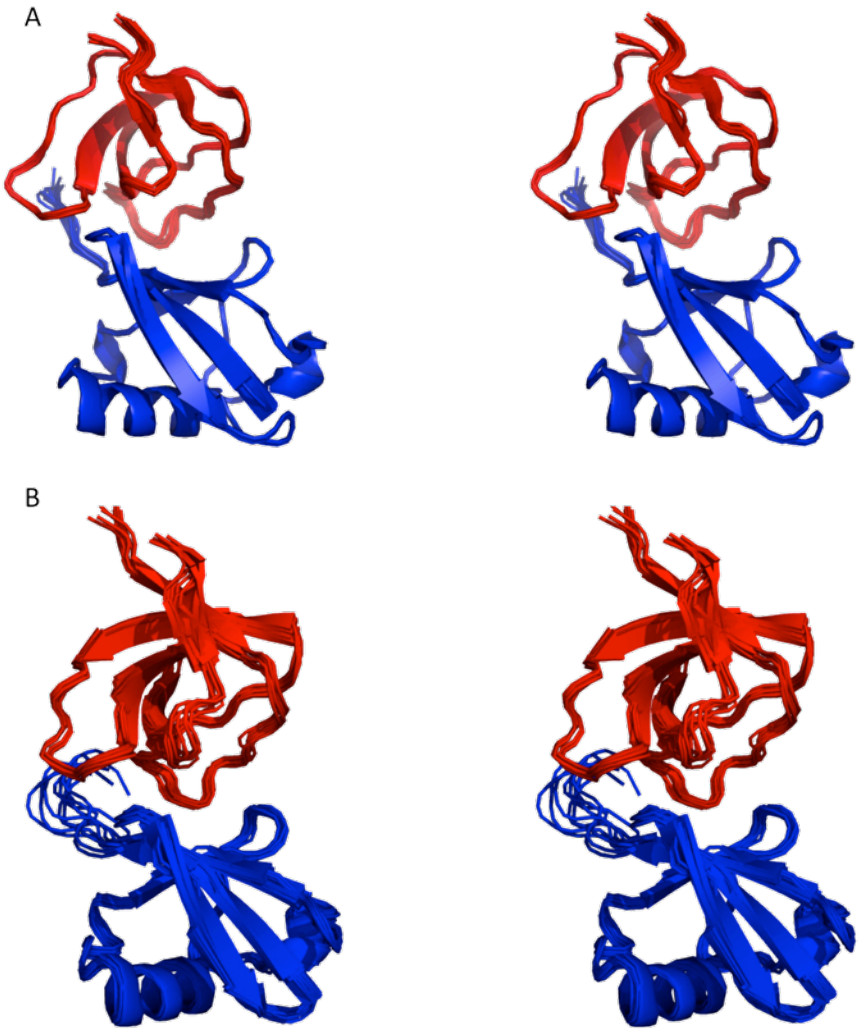


Figure 4: Solution structures of the complexes between ubiquitin and the first and third SH3 domains of CD2AP. *Stereo representation of the ensemble of 10 lowest combined target function structures of the CD2AP SH3-A (**A**) and C (**B**) domains in complex with ubiquitin. Ubiquitin is represented in blue and the SH3 domain in red. Structures were superimposed on the backbone atoms of residues 7-70 in ubiquitin.*

Solution structure of the SH3-C:Ubiquitin complex

We have recently described a novel approach for the structure determination of macromolecular complexes [12] that was applied to obtain the structure of the SH3-C:Ubiquitin complex. Figure 4B shows a stereo representation of the ensemble of the lowest target function structures (combining RDC, AIR and NOe violation) of CD2AP SH3-C in complex with ubiquitin. The ensemble shows very good Ramachandran statistics while fulfilling the experimental data (table 1), with an average backbone RMSD for residues 8-62 in SH3-C and 6-70 in ubiquitin of 1.90 Å.

Table 1: *Structural statistics for the 10 lowest combined target function structures of CD2AP SH3 A and C domains in complex with ubiquitin. Flexible N- and C-terminal residues were omitted from the RMSD analysis. Ramachandran statistics were obtained from PROCHECK-NMR analysis over all residues and RDC violations were calculated with MODULE [31].*

	SH3-A:Ubiquitin	**SH3-C:Ubiquitin**
Distance restraints		
AIRS	26	48
Orientational restraints		
Residual dipolar couplings	56	227
Restraint statistics		
NOe violations > 0.5 Å	-	0
H_N-N and C_α-H_α RDC violations > 3 Hz	2	7
H_N-C′ RDC violations > 1 Hz	-	2
C_α-C′ RDC violations > 0.5 Hz	-	3
RMSD from average (Å)		
Backbone N, CA, C′	1.42	1.90
Ramachandran plot		
Most favored regions (%)	78.1	87.0
Additional allowed regions (%)	20.5	10.4
Generously allowed regions (%)	1.4	2.6
Disallowed regions (%)	0.0	0.0

Solution structure of the SH3-A:Ubiquitin complex

Bound-form NH RDCs measured in a PEG-Hexanol mixture were introduced into a structure refinement procedure using the program SCULPTOR [23] starting from conformers from different clusters sampled

by HADDOCK [30] using only chemical shift perturbation as intermolecular restraint. The stereo representation of the ensemble of lowest target function structures (combining RDC and AIR violations) of CD2AP SH3-A in complex with ubiquitin is shown in figure 4A. Structural statistics are summarized in Table 1. The average backbone RMSD for residues 3-56 in SH3-A and 6-70 in ubiquitin is 1.42 Å.

Comparison between the different SH3:Ubiquitin complexes

We used the secondary structure matching algorithm as implemented in [32] to perform structural comparison between the different SH3:ubiquitin complexes. The structure of the SH3-A:Ubiquitin complex is more similar to the previously reported CIN85 SH3-C:Ubiquitin structure (PDB entry 2K6D) with a RMSD of 1.44 Å for the lowest-energy structures and in lesser extent to the Sla1 SH3-3:Ubiquitin structure (PDB entry 2JT4), with a RMSD of 1.55 Å. The structure of SH3-C:Ubiquitin complex is importantly different from the Sla1 SH3-3 and CIN85 SH3-C ubiquitin complexes with RMSDs of 1.92 Å to both structures. The total surface area that becomes buried upon complex formation is 1605.1 and 1166.4 Å for CD2AP SH3-A and C domains, respectively, in comparison to 1246.9 Å for Sla1 SH3-3 complex and 1667.0 Å for CIN85 SH3-C complex.

Figure 5 shows the residues directly involved in the intermolecular interactions between both CD2AP SH3 domains and ubiquitin. Both SH3 domains bind to the same face of ubiquitin, which involves the Ile/Leu-rich patch, the region of the protein that is most commonly found in complex with UBDs [9, 10], but differs mainly in the type of residues surrounding this hydrophobic patch. CD2AP SH3-A (Figure 6A) domain adopts a similar orientation than CIN85 SH3-C and Sla1 SH3-3 complexes with ubiquitin (Figure 6C and D), establishing non-bonded hydrophobic interactions between the key SH3 Phe53 and His68 of ubiquitin located in the Ile44 binding region.

Despite of the similar orientation, different contacts are found between ubiquitin and the RT and n-Src loops (Figure 5B). While in the CD2AP SH3-A:ubiquitin complex the RT loop engages Arg72, Arg74 and Gly75 and 76, located in the C-terminus of ubiquitin via hydrogen bonds and salt bridges with the polar and charged residues His14 and Asp15 and 16, Sla1 SH3-3 RT loop establishes non-bonding hydrophobic interactions only with the Ile44 binding region of ubiquitin via the same aromatic residues than CIN85 SH3-C. No major differences between these three SH3 domains are found in the interactions involving the n-Src loop with the Leu8 and the Val70 binding regions of ubiquitin, and the 3^{10} helix with the Leu8 and Ile44 binding regions. In the three cases Trp37 (391 in Sla1 SH3-3 and 306 in CIN85 SH3-C) targets the Leu8 and Val70 binding regions, and Phe53

(409 in Sla1 SH3-3 and 322 in CIN85 SH3-C) interacts with the Ile44 binding region mainly by hydrophobic interactions with His68 and/or Ile44.

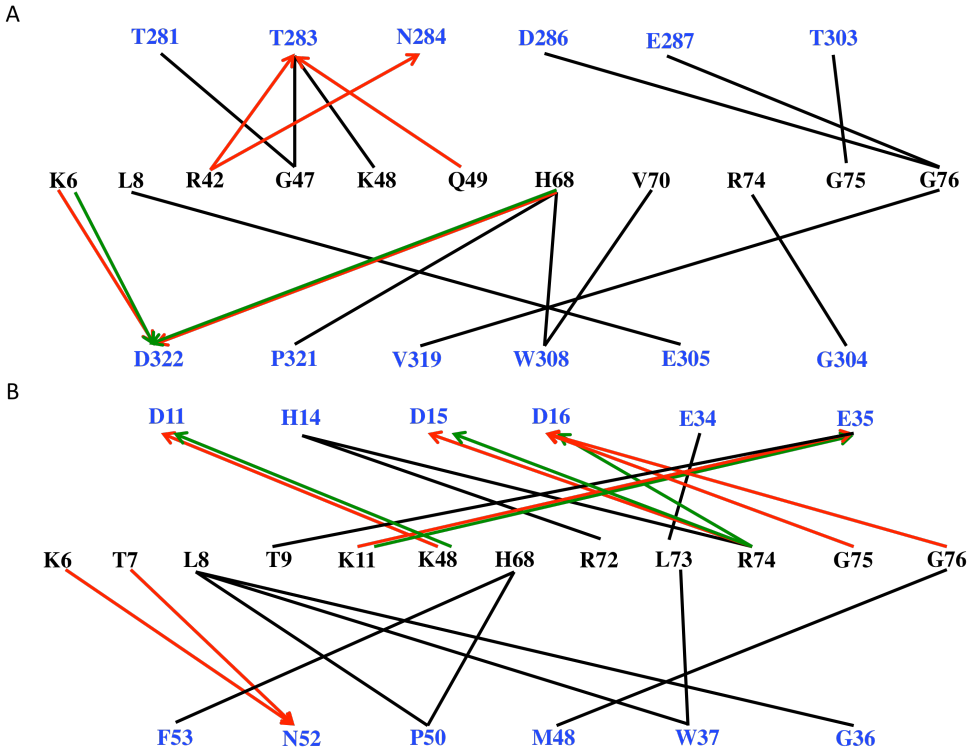


Figure 5: Intermolecular interactions detected in the complexes between CD2AP SH3-A and -C with ubiquitin. *Interactions are specified when detected in ≥ 50% of the 10 lowest target function structures of CD2AP SH3-C* ***(A)*** *and CD2AP SH3-A* ***(B)*** *in complex with ubiquitin. Ubiquitin residues are denoted in black. The lines connect interacting residues; black lines indicate hydrophobic contacts, red lines indicate hydrogen bonding interactions and green lines indicate salt bridges.*

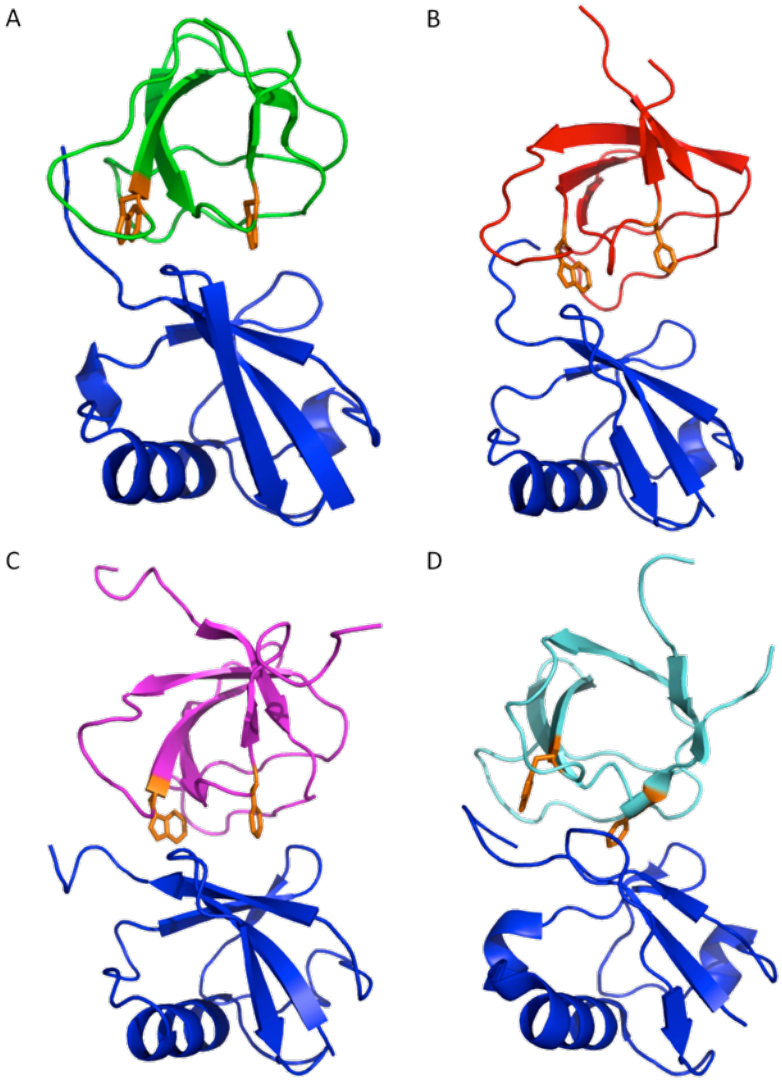


Figure 6: Structural comparison of different SH3:Ubiquitin complexes. *Cartoon representation of the lowest target function or lowest energy structure of different SH3:Ubiquitin complexes.* ***A.*** *CD2AP SH3-A:Ubiquitin complex. Ubiquitin is represented in blue and CD2AP SH3-A in green.* ***B.*** *CD2AP SH3-C:Ubiquitin complex. Ubiquitin is represented in blue and CD2AP SH3-C in red.* ***C.*** *Sla1 SH3-3 in complex with ubiquitin (PDB entry 2JT4; ubiquitin in blue, Sla1 SH3-3 in magenta).* ***D.*** *CIN85 SH3-C in complex with ubiquitin (PDB entry 2K6D; ubiquitin in blue, CIN85 SH3-C in cyan). Phenylalanine and tryptophan residues equivalent to F53 and W41 in CD2AP SH3-A are given in stick representation and coloured in orange in all SH3 domains. All complexes were superimposed on the backbone atoms of residues 4-70 of ubiquitin.*

The binding interface for CD2AP SH3-C (Figure 6B) engages both the Ile/Leu-rich hydrophobic patch and the C-terminal diglycine patch on ubiquitin but in a different way than the above mentioned SH3 domains. This orientation is dominated by non-bonded hydrophobic interactions of Thr303 and Gly304 residues in the n-Src loop with Arg74 and the diglycine 75-76 patch in ubiquitin, plus an additional non bonded interaction of Val319 with Gly76, in agreement with the high chemical shift perturbation observed in the C-terminus of ubiquitin (Figure 2D). Furthermore, the RT loop also targets the Gly76 binding region via its polar residues Asp286 and Glu287, as well as the Ile44 binding region via hydrogen bonds with Thr283 and Asn284. Trp308 is also involved in hydrophobic interactions with Val70 and His68 in ubiquitin, but not with the Leu8 binding region. The 3^{10} helix targets His68 and Lys6 in ubiquitin through the formation of two hydrogen bonds, but importantly Phe324 is not involved in any contact with ubiquitin.

Mutational Analysis of the CD2AP SH3:Ubiquitin interactions

As stated before, ubiquitin-binding activity by SH3 domains has been related with the presence of a phenylalanine in equivalent positions to F409 in the Sla1 SH3-3 domain [13]. Close inspection of both CD2AP SH3:Ubiquitin complexes indicate that Phenylalanine residue 53 in CD2AP SH3-A (Figure 6A, coloured in orange) is located at the heart of the hydrophobic interface, similarly positioned than the equivalent phenylalanine in the CIN85 SH3-C and Sla1 SH3-3:ubiquitin complexes (Figure 6C and D). In contrast, phenylalanine residue 324 in CD2AP SH3-C (Figure 6B, coloured in orange) is not directly involved in binding as it is placed at the edge of the binding interface. Mutation of this phenylalanine to a tyrosine in the CD2AP SH3-C domain is thus not to be expected to abolish ubiquitin binding as was observed for Sla1 SH3-3. NMR spectra recorded on a ^{15}N-labeled ubiquitin sample with increasing amounts of the SH3-C F324Y mutant (Figure 7A) indicate that the mutant binds ubiquitin in a similar way as the WT protein, as expected from the three-dimensional structure. Equivalent experiments were performed with CD2AP SH3-A F53Y mutant. In this case the F53Y mutation in CD2AP SH3-A abolished binding as observed from the ^{1}H-^{15}N HSQC spectra (Figure 7B), in agreement with our structural model and with the mutational analysis performed on Sla1 SH3-3 [13].

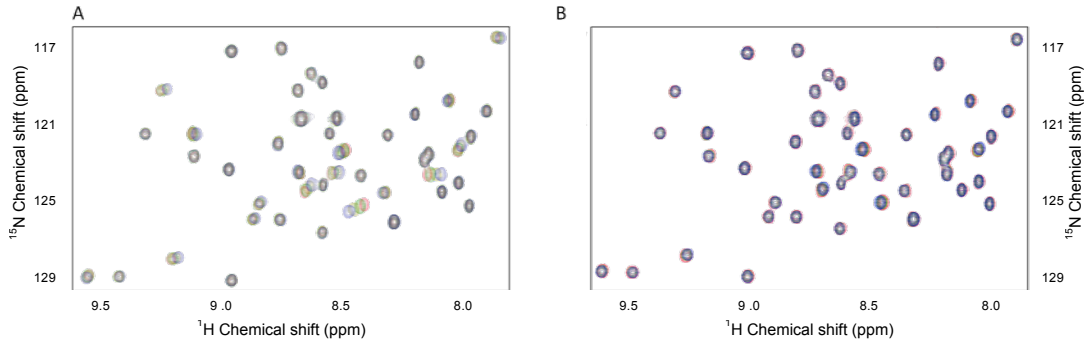


Figure 7: Mutational study of the ubiquitin binding by the CD2AP SH3-A and C domains. *Region of ^{1}H-^{15}N HSQC spectra recorded at increasing amounts of the CD2AP SH3 mutant (red to blue).* ***A.*** *^{1}H and ^{15}N Chemical shift changes upon titrating the CD2AP SH3-C F324Y mutant into a ^{15}N-labelled ubiquitin sample up to a ratio of 1:0.9 (Ubi:SH3).* ***B.*** *^{1}H and ^{15}N Chemical shift changes upon titrating the CD2AP SH3-A F53Y mutant into a ^{15}N-labelled ubiquitin sample up to a ratio of 1:1 (Ubi:SH3).*

To confirm our NMR studies we performed fluorescence and Isothermal Titration Calorimetry (ITC) experiments on WT CD2AP SH3-A and C and on the F324Y mutant of CD2AP SH3-C with ubiquitin (data not shown). Both SH3-C variants bind ubiquitin with a moderate but similar affinity with a K_d of 12 ± 2 μM and 24 ± 3 μM for the WT and the F324Y mutant, respectively, while SH3-A binds ubiquitin with substantial higher affinity with a K_d of 1.9 ± 0.3 μM (data not shown), in agreement with the intermediate exchange observed in the NMR titration experiments.

Effect of the extension of the ubiquitin C-terminus in SH3 binding.

Hicke and co-workers showed that some of the ubiquitin interacting SH3 domains bind significantly better to ubiquitin when GST was covalently attached to its C-terminus, including the CIN85 SH3-C domain and the amphiphysin SH3 domains [13]. We investigated if such behaviour is also found for the CD2AP SH3-A and SH3-C domains and if this is somehow related to the difference in mode of ubiquitin binding. Figure 8A shows NMR detected titrations of non-His tagged ubiquitin versus C-terminal His-tagged ubiquitin into ^{15}N-labelled SH3-C. One order of magnitude higher K_d is obtained for titration of SH3-C with non His-tagged ubiquitin, although no direct interaction of the His-tag is observed in the ^{1}H-^{15}N HSQC and no differences are found in the ubiquitin areas targeted by this SH3 domain (Figure 8B). In contrast, such a significant increase in affinity is not observed in ubiquitin binding by the CD2AP SH3-A domain (data not shown) and by Sla1 SH3-3 [13]. In light of these results, and due to the

higher CIN85 SH3-C affinity to C-terminal GST-tagged ubiquitin reported before [13], a NMR detected titration was carried out with unlabelled CIN85 SH3-C and ^{15}N-labelled C-terminal His-tagged ubiquitin. Interestingly, our NMR titrations suggest significantly tighter binding and different areas in ubiquitin are affected (Figure 3B) than those reported previously by Forman-Kay and co-workers using non-tagged ubiquitin [15]. Furthermore, the ubiquitin areas targeted described in this work are very similar to the CD2AP SH3-C titration. Thus CIN85 SH3-C is showing similar behaviour as CD2AP SH3-C towards ubiquitin binding.

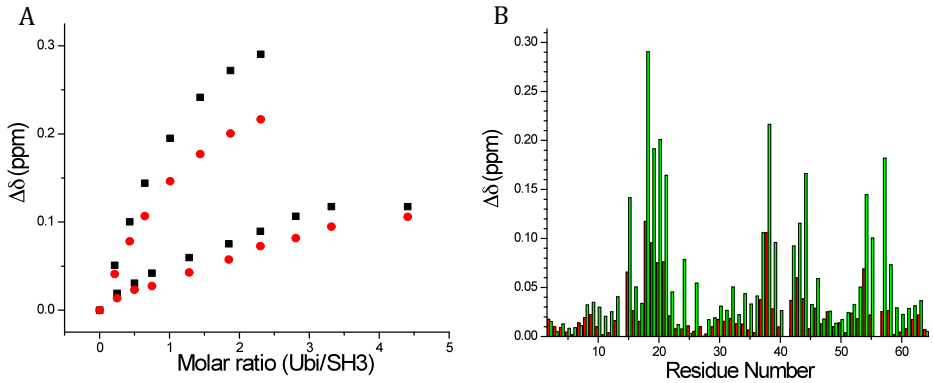


Figure 8: The effect of a C-terminal tag in Ubiquitin on SH3 binding. A. *Titration curve monitored by NMR chemical shift perturbations of Thr283 (black) and Glu303 (red) in CD2AP SH3-C. The upper curves correspond to titrations with C-terminal His-tagged ubiquitin, while the lower curves correspond to titrations with non-tagged ubiquitin.* ***B.*** *Comparison of the chemical shift deviations (Δδ) between the first and last titration point of titrating ^{15}N-labelled CD2AP SH3-C with C-terminal His-tagged ubiquitin (green) and non-tagged ubiquitin (red).*

DISCUSSION

Ubiquitin is known to regulate a wide variety of cellular activities ranging from transcriptional regulation to cell signalling and membrane trafficking. Many cellular activities of ubiquitin are mediated by mono- rather than poly-ubiquitin, and its functions are deciphered by various ubiquitin-binding proteins. Recently it was found that a subset of SH3 domains constitute a new, distinct type of ubiquitin-binding domains [13]. The structure of the complex between Sla1 SH3-3 and ubiquitin shows that the ubiquitin binding surface overlaps largely with the canonical binding surface for proline-rich ligands and that like many other ubiquitin-binding motifs, the SH3 domain engages the Ile44 hydrophobic patch of ubiquitin [14]. A key affinity and specificity determinant for ubiquitin-binding was appointed to Phe409 of Sla1, located at the heart of the hydrophobic interface in the SH3-ubiquitin complex. Moreover, a tyrosine residue at the equivalent position was found to abrogate ubiquitin binding.

We previously used NMR residual dipolar couplings to obtain the solution structure of the complex between ubiquitin and the third (SH3-C) domain of CD2AP [12]. RDCs have been shown to provide orientational information crucial for the determination of an accurate conformation of macromolecular complexes in conditions where the measurement of intermolecular NOe might be compromised due to the insensitivity of NOESY experiments. In this work, we use this approach to obtain the solution structure of the complex between ubiquitin and the first (SH3-A) domain of CD2AP. Mutation of Phe53 (equivalent to Phe409 in Sla1 SH3-3) to a tyrosine abolishes ubiquitin binding similar to what was observed for Sla1 SH3-3. In contrast mutation of the corresponding phenylalanine residue in CD2AP SH3-C has a minor effect on ubiquitin binding as observed in both the NMR titration and ITC experiments. The results obtained here demonstrate that the two SH3 domains from the same adaptor protein show alternative modes of ubiquitin binding. The intriguing question that remains is which residues and/or what region of a particular ubiquitin-binding SH3 domain determines the mode/orientation of binding to ubiquitin and if this difference in binding mode is somehow related to a difference in function.

The distinct interactions between SH3 domains and ubiquitin can only be rationalized analyzing the surfaces of both proteins. Figure 9C shows that the CD2AP SH3C binding surface is completely negatively charged, in contrast with the positive and non-polar areas observed in CD2AP SH3-A (Figure 9A) and Sla1 SH3-3. Close inspection of the interactions involved in the different SH3:Ubiquitin structures indicate that the main differences between both binding modes rely in the interactions between polar residues in the RT and n-Src loops and the ubiquitin C-terminus.

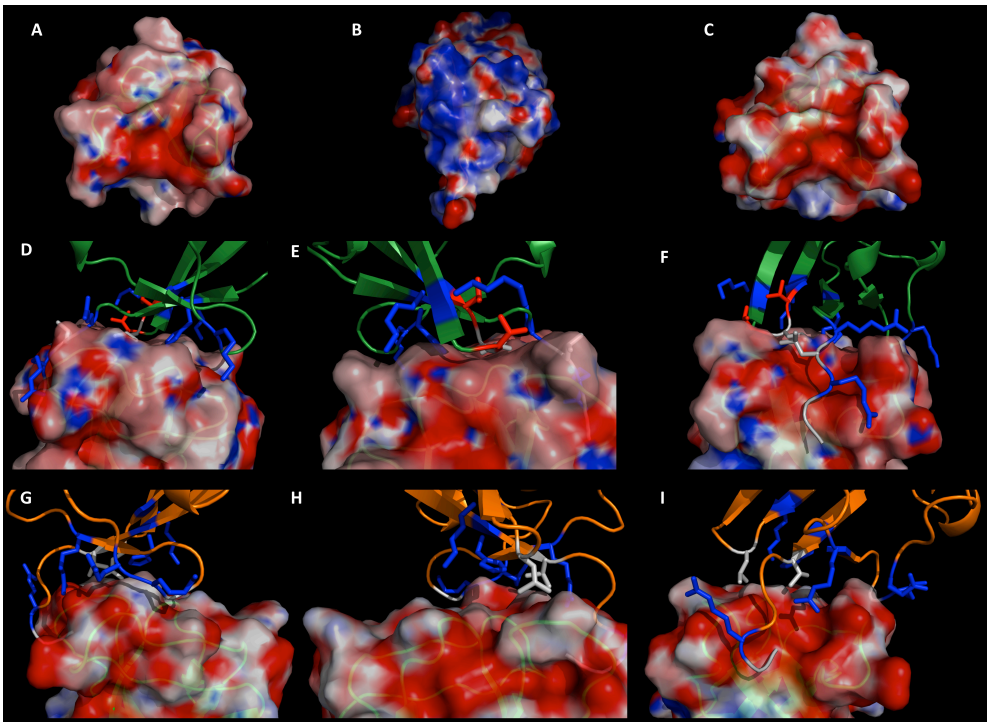


Figure 9. SH3:Ubiquitin binding interfaces. *Surface representations of the free forms of CD2AP SH3-A (**A**), ubiquitin (**B**) and CD2AP SH3-C (**C**), showing the surface-charge distribution. Electrostatic surface representation of the orientation of the CD2AP SH3-A and C complexes with ubiquitin (cyan and orange respectively), zooming in at the SH3 RT-loop (**D.** and **G** respectively), SH3 n-Src loop (**E** and **H** respectively) and ubiquitin C-terminus (**F** and **I** respectively). Ubiquitin residues involved in intermolecular contacts are represented in coloured sticks (grey for non polar residues, blue and red for positively and negatively charged or polar residues respectively). The electrostatic surface representation was drawn with Pymol (www.pymol.org) using a Poisson-Boltzmann electrostatics calculation [33].*

Comparison of the amino acid sequences of CD2AP SH3-A and C, Sla1 SH3-3 and CIN85 SH3-C domains (Figure 11) displays that the residues found in the RT loop are generally similar, except for the presence of two negatively polar residues in positions 281 and 283 in CD2AP SH3-C that are replaced by a negatively charged and a hydrophobic residue in CD2AP SH3-A and Sla1 SH3-3 domains. This results in favourable additional interactions between this area and Arg42 in ubiquitin and a consequent displacement of the interface that is not observed in CD2AP SH3-A and Sla1 SH3-3 ubiquitin complexes. Bigger differences are observed in the n-Src loop. This loop in CD2AP SH3-C is one residue longer and secondly exposes a completely negative charged surface, allowing a closer disposition of the ubiquitin C-terminus towards it and thus non-bonded interactions involving Arg74 and Gly75 in ubiquitin with Thr303 and

Gly304. CD2AP SH3-A and Sla1 SH3-3 present a shorter n-Src loop with a mixture of negatively and positively charged residues in the surface, hindering any interaction with the C-terminus of ubiquitin.

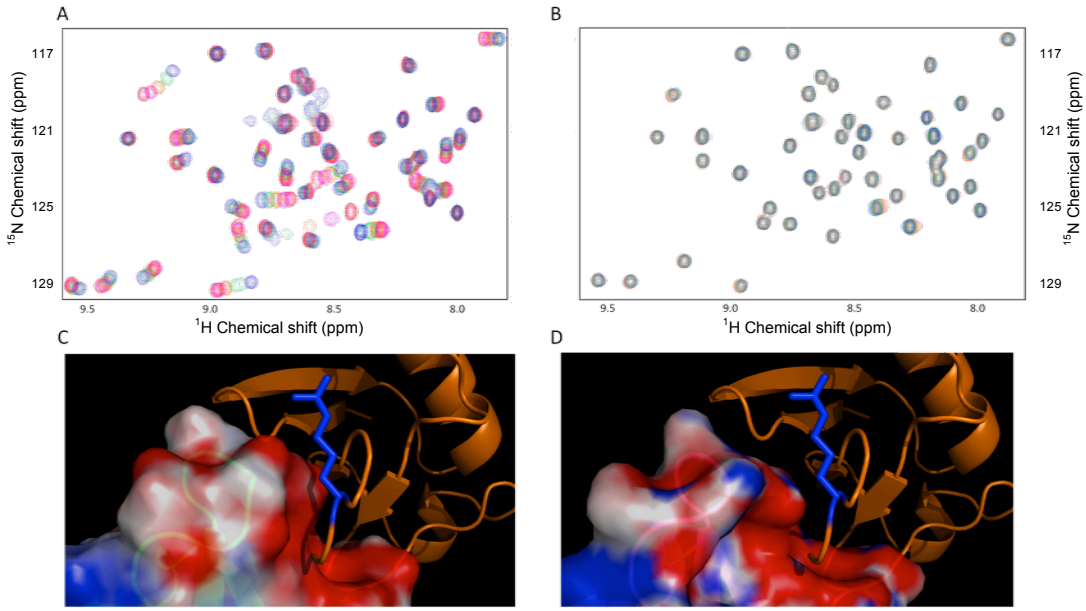


Figure 10. Mutational study of the ubiquitin binding by the CD2AP SH3-C domain. *Region of ^{1}H-^{15}N HSQC spectra recorded at increasing amounts of the CD2AP SH3 mutant (red to blue).* ***A.*** *^{1}H and ^{15}N Chemical shift changes upon titrating the CD2AP SH3-C T283A mutant into a ^{15}N-labelled ubiquitin sample up to a ratio of 1:1.4 (Ubi:SH3).* ***B.*** *^{1}H and ^{15}N Chemical shift changes upon titrating the CD2AP SH3-C E302K + T303 deletion mutant into a ^{15}N-labelled ubiquitin sample up to a ratio of 1:1.1 (Ubi:SH3).* ***C*** *and* ***D.*** *Electrostatic surface representation of the interaction between the ubiquitin C-terminus and the n-Src loop of the WT CD2AP SH3-C and the CD2AP SH3-C E302 + T303 deletion mutant respectively. Ubiquitin Arg74 side chain is represented in blue sticks.*

To rationalize the importance of the n-Src loop over the RT loop in this case we constructed two SH3-C mutants, one in which Thr283 in the RT loop was mutated into an alanine, and one in which Glu302 in the n-Src loop was mutated into a lysine together with the deletion of Thr303. The latter mutant thus becomes more similar to the Sla1 SH3-3 domain in terms of charge and length of the n-Src loop. No significant difference in ubiquitin binding was found between the WT and the Thr283Ala mutant (Figure 10A), similarly to what was found for the glutamate to alanine mutation at the equivalent position in Sla1 SH3-3 [13]. In contrast, the Glu302Lys + Thr303 deletion mutant completely eliminates ubiquitin binding (Figure 10B), thus pointing to the n-Src loop as the key region involved in this ubiquitin binding mode. To discard that deficiency of ubiquitin binding resulted from an unfolded SH3-C domain due to the

mutation, 1D ^{1}H experiments were recorded throughout the NMR titration experiments. These spectra indicate proper folding of the SH3-C mutant (data not shown).

We constructed a structural model of the CD2AP SH3-C E302K + T303 deletion mutant using the program Modeller [34]. The shorter n-Src loop moves this region further away from the ubiquitin C-terminus (Figure 10D). At the same time, the replacement of the Glu302 for a Lys changes the net charge in the surface of the loop, thus eliminating the interaction of Thr303 with Gly75 and hindering the interaction of Gly304 with the Arg74 in ubiquitin. As stated before, positively charged and non-polar residues constitute the same regions in CD2AP SH3-A and in Sla1 SH3-3. These residues are not involved in any intermolecular interaction in the ubiquitin complexes. Both differences in length and charge of the n-Src loop are thus related to the different ubiquitin binding modes exhibited by SH3 domains. All mutations performed in ubiquitin binding SH3 domains and their effects are summarized in Figure 11.

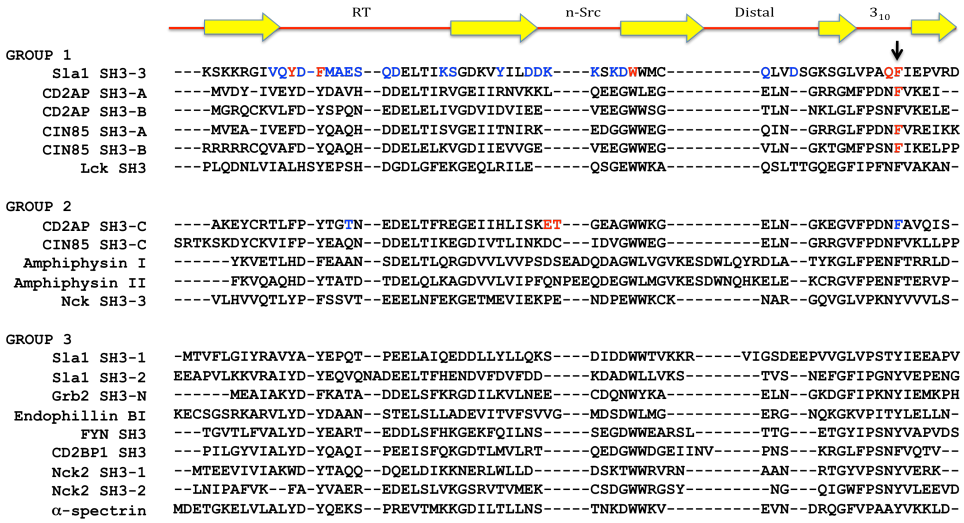


Figure 11. Sequence comparison between ubiquitin-binding SH3 domains. *Sequence alignment of different types of SH3 domains: Group 1 is constituted by SH3 domains binding ubiquitin in a similar orientation than Sla1 SH3-3. Group 2 is constituted by SH3 domains binding ubiquitin in a similar orientation than CD2AP SH3-C. Group 3 is constituted by SH3 domains not able to bind ubiquitin. Mutations abolishing ubiquitin binding are highlighted in red, and mutations with no considerable effect are highlighted in blue [13, 15]. The black arrow points at the position of the suggested key phenylalanine [13]. The secondary structure elements are shown at the top of the figure.*

Hicke and co-workers showed that some of the ubiquitin interacting SH3 domains bind significantly better to ubiquitin when GST was covalently

attached to its C-terminus, including the CIN85 SH3-C domain and the amphiphysin SH3 domains [13]. The importance of the ubiquitin C-terminus is not novel in SH3:ubiquitin binding as observed in the interaction between Endophilin-A SH3 and the Parkin ubiquitin like protein (Ubl) [35] and in the interaction between NcK SH3-3 and ubiquitin [11]. Although no structure of the complex with ubiquitin was reported, the authors showed that Nck SH3-3 domain interacts with ubiquitin even though it posses a tyrosine residue instead of a phenylalanine at the key position 73. Interestingly chemical shift perturbations (CSP) suggest that the ubiquitin C-terminus is directly involved in SH3 binding, although different CSP are found in ubiquitin and the Nck SH3-3 domains than in the Sla1 SH3-3 and CD2AP SH3-C ubiquitin binding modes. In the case of the interaction between Endophilin-A SH3 and the Parkin Ubl the orientation adopted of both partners is considerably different than that of Sla1 SH3-3 or CD2AP SH3-C. The Parkin Ubl C-terminus adopts an extended β strand conformation that mimics a proline-rich ligand. Moreover, Arg75 mutation to a Gly residue (the corresponding residue at position 75 in ubiquitin) completely abolishes binding, whereas Lys76 to Gly mutation diminishes the interaction [19]. Furthermore, the Endophilin-A SH3 domain is considerably shifted towards the C-terminus compared with the CD2AP SH3-C domain. Therefore, CD2AP SH3-C:ubiquitin interaction constitutes a different ubiquitin binding mode than those described for Sla1 SH3-3 domain and for Endophilin-A SH3 domains.

Interestingly, CIN85 SH3-C titrations with ^{15}N-labelled C-terminal His-tagged ubiquitin (Figure 3B) revealed a chemical shift perturbation pattern very similar to the one exhibited by CD2AP SH3-C (Figure 2D), inconsistent with the structural model between the CIN85 SH3-C domain and non-tagged ubiquitin obtained by Forman-Kay and co-workers [15]. According to our NMR titration, main chemical shift differences upon binding are located in the ubiquitin C-terminal region and in the Ile44 binding region. In contrast, the Leu8 binding region displays much lower chemical shift differences. Interestingly, in another work it was suggested that CIN85 SH3-C presents two distinct conformations when bound to ubiquitin [36]. The first binding mode, originally named $(PL)_1$, presents a lower network of contacts involving Leu8 and Ile44. In contrast, His68 and Gly47-Gln49 establish contacts with the RT loop, as well as the C-terminal region of ubiquitin, in perfect agreement with the chemical shift perturbation pattern found in CIN85 SH3-C binding to C-terminal tagged ubiquitin. Moreover, no contact involving the key Phe322 is observed, suggesting that CIN85 SH3-C is able to bind ubiquitin in both the canonical (Sla1 SH3-3) and the CD2AP SH3-C mode. As observed from the sequence alignment (Figure 11), CD2AP SH3-C and CIN85 SH3-C share similar charge and length in the n-Src loop, what would confirm the importance of this loop for this ubiquitin binding mode. Forman-Kay and co-workers showed that the triple Phe-to-Tyr mutant in full-length CIN85 is still able to

ubiquitinate CIN85 through interaction with the Cbl E3 ubiquitin protein ligase like the WT protein [15]. They suggested that in this case none of the three CIN85 SH3 domains is able to grab ubiquitin and that even in the absence of EGF stimulation and consequent Cbl inactivation CIN85 is ubiquitinated. The possibility of CIN85 SH3-C to bind ubiquitin as CD2AP SH3-C explains that even though having a tyrosine at position 322 the triple mutant CIN85 is still capable of binding ubiquitin via its SH3-C domain suggesting the same mechanism of ubiquitination as for the WT protein.

The presence of a C-terminal tag in ubiquitin has a strong effect in the affinity ubiquitin for CD2AP SH3-C for but not for CD2AP SH3-A domain, even though the tag is not involved in the interaction. The extension of the C-terminus might result in rigidification of the ubiquitin C-terminus, by lowering the entropy cost associated with the decrease of its flexibility upon binding, as observed in [35], what would be reflected in an increase in affinity. This effect was previously reported [13] for CIN85 SH3-C and amphiphysin SH3 domains binding to ubiquitin, although in this case the His-tag was replaced by a GST-tag. These domains, like CD2AP SH3-C, posses a longer n-Src loop, as well as negatively charged residues at the edge of it, suggesting that they bind ubiquitin similar to CD2AP SH3-C when binding C-terminal tagged ubiquitin. This suggests that the higher affinity to C-terminal tagged ubiquitin constitute a feature of this binding mode.

CD2AP SH3-A and B domains are able to tightly bind free ubiquitin. For CD2AP SH3-C, however, tight binding requires that the ubiquitin molecule has to be attached via its C-terminus to a protein, avoiding in this way the interaction of this domain with free ubiquitin present in the cellular milieu. It has been described that CIN85/CD2AP are ubiquitinated through interaction with the Cbl E3 ubiquitin protein ligase mediated by interaction of their SH3-A and/or SH3-B domains with the proline-rich domain of c-Cbl upon EGF activation [37, 38]. The ubiquitin molecule covalently attached to CIN85/CD2AP could be then an ideal candidate for the interaction of the SH3-C domain.

We conclude that CD2AP SH3-C domain interaction with ubiquitin constitutes a new ubiquitin binding mode involved in a different cellular function, and thus changes the previously established mechanism of EGF-dependent CD2AP/CIN85 mono-ubiquitination.

ACKNOWLEDGEMENTS

This research was funded by grant BIO2005-04650 from the Spanish Ministry of Education and Science (MEC) and FQM-02838 from the Andalucia Regional Goverment. We thank Francisco Conejero-Lara for

useful discussions and Obdulio Lopez-Mayorga for his continuous support. J.L.O.R. is supported by a FPU pre-doctoral grant from the MEC. S.C.A. is recipient of a return grant of the University of Granada. The CD2AP SH3-A clone was kindly provided by Jeronimo Bravo. The 600 MHz spectra were recorded in the Centre for Scientific Instrumentation (CIC) of the University of Granada and at the RALF Large Scale Facility in Grenoble, which is funded by the 'Access to Research Infrastructures' program of the European Union.

REFERENCES

1. Pawson, T., M. Raina, and P. Nash, *Interaction domains: from simple binding events to complex cellular behavior.* FEBS Lett, 2002. **513**(1): p. 2-10.
2. Schlessinger, J., *Cell signaling by receptor tyrosine kinases.* Cell, 2000. **103**(2): p. 211-25.
3. Waterman, H. and Y. Yarden, *Molecular mechanisms underlying endocytosis and sorting of ErbB receptor tyrosine kinases.* FEBS Lett, 2001. **490**(3): p. 142-52.
4. Kirsch, K.H., et al., *The adapter type protein CMS/CD2AP binds to the proto-oncogenic protein c-Cbl through a tyrosine phosphorylation-regulated Src homology 3 domain interaction.* J Biol Chem, 2001. **276**(7): p. 4957-63.
5. Petrelli, A., et al., *The endophilin-CIN85-Cbl complex mediates ligand-dependent downregulation of c-Met.* Nature, 2002. **416**(6877): p. 187-90.
6. Soubeyran, P., et al., *Cbl-CIN85-endophilin complex mediates ligand-induced downregulation of EGF receptors.* Nature, 2002. **416**(6877): p. 183-7.
7. Szymkiewicz, I., et al., *CIN85 participates in Cbl-b-mediated down-regulation of receptor tyrosine kinases.* J Biol Chem, 2002. **277**(42): p. 39666-72.
8. Dikic, I., *CIN85/CMS family of adaptor molecules.* FEBS Lett, 2002. **529**(1): p. 110-5.
9. Hicke, L., H.L. Schubert, and C.P. Hill, *Ubiquitin-binding domains.* Nat Rev Mol Cell Biol, 2005. **6**(8): p. 610-21.
10. Hurley, J.H., S. Lee, and G. Prag, *Ubiquitin-binding domains.* Biochem J, 2006. **399**(3): p. 361-72.
11. Kang, J., et al., *Distinct interactions between ubiquitin and the SH3 domains involved in immune signaling.* Biochim Biophys Acta, 2008. **1784**(9): p. 1335-41.
12. Ortega-Roldan, J.L., et al., *Accurate characterization of weak macromolecular interactions by titration of NMR residual dipolar*

couplings: application to the CD2AP SH3-C:ubiquitin complex. Nucleic Acids Res, 2009. **37**(9): p. e70.

13. Stamenova, S.D., et al., *Ubiquitin binds to and regulates a subset of SH3 domains.* Mol Cell, 2007. **25**(2): p. 273-84.
14. He, Y., L. Hicke, and I. Radhakrishnan, *Structural basis for ubiquitin recognition by SH3 domains.* J Mol Biol, 2007. **373**(1): p. 190-6.
15. Bezsonova, I., et al., *Interactions between the three CIN85 SH3 domains and ubiquitin: implications for CIN85 ubiquitination.* Biochemistry, 2008. **47**(34): p. 8937-49.
16. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.
17. Johnson, B.A., *Using NMRView to visualize and analyze the NMR spectra of macromolecules.* Methods Mol Biol, 2004. **278**: p. 313-52.
18. Goddart, T.D.a.K., D.G., University of California, 1993.
19. Ruckert, M. and G. Otting, *Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments.* Journal of the American Chemical Society, 2000. **122**(32): p. 7793-7797.
20. Brutscher, B., *Accurate measurement of small spin-spin couplings in partially aligned molecules using a novel J-mismatch compensated spin-state-selection filter.* J Magn Reson, 2001. **151**(2): p. 332-8.
21. Lescop, E., P. Schanda, and B. Brutscher, *A set of BEST triple-resonance experiments for time-optimized protein resonance assignment.* J Magn Reson, 2007. **187**(1): p. 163-9.
22. Schanda, P., H. Van Melckebeke, and B. Brutscher, *Speeding up three-dimensional protein NMR experiments to a few minutes.* J Am Chem Soc, 2006. **128**(28): p. 9042-3.
23. Hus, J.C., D. Marion, and M. Blackledge, *De novo determination of protein structure by NMR using orientational and long-range order restraints.* J Mol Biol, 2000. **298**(5): p. 927-36.
24. Morris, A.L., et al., *Stereochemical quality of protein structure coordinates.* Proteins, 1992. **12**(4): p. 345-64.
25. Laskowski, R.A., et al., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* J Biomol NMR, 1996. **8**(4): p. 477-86.
26. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.* Protein Eng, 1995. **8**(2): p. 127-34.
27. E. Krissinel, K.H. *Protein interfaces, surfaces and assemblies service PISA at the European Bioinformatics Institute*. Available from: http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html.

28. Casares, S., et al., *Cooperative propagation of local stability changes from low-stability and high-stability regions in a SH3 domain.* Proteins, 2007. **67**(3): p. 531-47.
29. Zuiderweg, E.R., *Mapping protein-protein interactions in solution by NMR spectroscopy.* Biochemistry, 2002. **41**(1): p. 1-7.
30. Dominguez, C., R. Boelens, and A.M. Bonvin, *HADDOCK: a protein-protein docking approach based on biochemical or biophysical information.* J Am Chem Soc, 2003. **125**(7): p. 1731-7.
31. Dosset, P., et al., *A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings.* J Biomol NMR, 2001. **20**(3): p. 223-31.
32. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
33. Dolinsky, T.J., et al., *PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W665-7.
34. Eswar, N., et al., *Protein structure modeling with MODELLER.* Methods Mol Biol, 2008. **426**: p. 145-59.
35. Trempe, J.F., et al., *SH3 domains from a subset of BAR proteins define a Ubl-binding domain and implicate parkin in synaptic ubiquitination.* Mol Cell, 2009. **36**(6): p. 1034-47.
36. Korzhnev, D.M., et al., *Alternate binding modes for a ubiquitin-SH3 domain interaction studied by NMR spectroscopy.* J Mol Biol, 2009. **386**(2): p. 391-405.
37. Haglund, K., et al., *Cbl-directed monoubiquitination of CIN85 is involved in regulation of ligand-induced degradation of EGF receptors.* Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12191-6.
38. Verdier, F., et al., *Ruk is ubiquitinated but not degraded by the proteasome.* Eur J Biochem, 2002. **269**(14): p. 3402-8.

Chapter 6

Final discussion

The CIN85/CD2AP family of adaptor proteins act as anchoring proteins that are involved in various signalling cascades [1, 2]. Several of the interactions in which they are involved are mediated via its three N-terminal SH3 domains. These SH3 domains are highly homologous amongst themselves. Sequences alignment shows highest similarity between the SH3-A and C domains, with 53.2% of identity, followed by a 51% identity between SH3-B and C and a 46.2% identity between SH3-A and B. However, it is still unknown what is the exact function of each CD2AP SH3 domain inside the cell, as their high sequence homology might be translated into overlapping specificities in binding to natural targets. Indeed it is known that these SH3 domains are able to interact, among others, with CD2 [1, 3], ALIX [4], c-Cbl [5, 6] and ubiquitin [7, 8]. The aim of this thesis consists in detailed structural and thermodynamic characterization of the three SH3 domains of CD2AP, as well as their interactions with ubiquitin. This thesis is part of a bigger project that aims to understand the determinants of the interaction between the different CD2AP SH3 domains with their natural targets and the effect of the presence of each individual domain on their neighbouring domain as it happens in the whole protein.

One of the obligatory steps to carry out our study consists in the cloning, expression and purification of the isolated CD2AP SH3 domains. This was done in both labelled and unlabelled media. Only in the case of SH3-B we experienced some problems related to its low stability and aggregation propensity, limiting its use for NMR.

To understand how a protein carries out its biological function it is necessary to know its three dimensional structure at atomic level. In this thesis we have used NMR for structural studies, as well as identification of binding sites, to drive the docking, and dynamic studies by ^{15}N relaxation, H/D exchange and RDCs. All these applications require the chemical shift assignment of at least the backbone of each domain. The availability of ^{13}C,^{15}N enriched SH3-A and C has allowed us to use a set of triple resonance spectra for the assignment of the backbone and side chain proton, carbon and nitrogen frequencies. The backbone assignment was based on the ^{15}N HSQC based spectra HNCACB, CBCA(CO)NH and HNCO, using a modified version of SmartNotebook [9], a novel semi-automatic assignment program that allowed the assignment of all backbone frequencies in a few hours. Hence, we have used this approach for the assignment of all peaks in the ^{1}H-^{15}N HSQC at different pH values and for their complexes. The assignment of SH3-C side chain protons and carbon frequencies was performed using HBHA(CO)NH, HCCH-TOCSY, ^{15}N-edited 3D TOCSY and 3D NOESY spectra, while the aromatic protons Hε and H_{δ} frequencies were obtained from a combined use of the (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE spectra. This allowed us to carry out the structure determination of the high-resolution structure of the SH3-C domain using a traditional NOe based structure calculation

protocol, as described in Chapter 2. The three-dimensional structure of CD2AP SH3-C shows the 5 β-strands and the 3^{10} helix that are common in the family of SH3 domains. It also contains all the features that are typically found in other SH3 domains, including the general binding site for the recognition of polyproline sequences that consists of a hydrophobic surface with three shallow pockets defined by the side chains of the preserved aromatic residues in SH3 domains. Due to the unavailability of isotopically enriched samples, the complete assignment of all ^{1}H in CD2AP SH3-B was carried out using a traditional homo-nuclear approach [10], combining ^{1}H-^{1}H 2D TOCSY and 2D NOESY spectra.

A detailed comparison of the three domains, including thermodynamics and backbone dynamics and their structural determinants required a set of high resolution structures in the same experimental conditions, as previous studies on SH3 domains have described large differences in stability and backbone dynamics upon changing the pH or salt concentration [11]. Consequently, in chapter 3 we have determined the solution structures of CD2AP SH3-B and C and refined the X-Ray structure of SH3-A [3] under exactly the same experimental conditions (pH 6.0). Comparison of the three structures revealed a high similarity between the three, except for the n-Src loop, which is different in length in the three domains. Moreover, the aromatic residues that are involved in recognition of polyproline sequences are very similarly positioned. This suggests that their binding specificity must come from small differences at specific sites. A similar electrostatic distribution is found on the surface around the polyproline binding pocket in SH3-A and B domains, dominated by a large negative area in the specificity pocket and a non-polar surface in the hydrophobic pockets surrounded by positive and negatively charged residues. In contrast, SH3-C presents a higher negatively charged surface around the hydrophobic pockets. The differences in charge distribution on the surface of the three domains might be related with the different binding affinities and is discussed below.

The biological function of the protein is also related to its stability. It is important then to get a complete picture of the stability of each of the domains. To do so a complete thermodynamic study was carried out by differential scanning calorimetry (DSC) (chapter 3). We have obtained the global unfolding data by DSC at different pH values for the three domains, and compared these among them and with other SH3 domains. We found that the SH3-A domain presents a higher stability than SH3-B and C. Interestingly, while these two domains show similar ΔG than other non-related SH3 domains, such as those of Abl, Fyn, Itk, Btk and Tec [12, 13], SH3-A shows an average 0.2 kJ/res higher experimental ΔG of unfolding (ΔG_U). The differences in stability were related with the structure using a computational structure based energy calculation using the program FOLDX [14], which provides the contribution of the intramolecular

interactions to the ΔG_U. FOLDX predicts an average 1.07 kJ/res higher enthalpy value for SH3-A driven by a higher number of hydrogen bonds and van der Waals interactions. Nevertheless, SH3-B also displays a similar contribution of the hydrogen bonds to the total free energy, but it cannot be related with the experimental enthalpy value probably due to the presence of high molecular weight species observed at room temperature by DLS.

The global folding data obtained by DSC was compared with the residue level information provided by NMR detected amide Hydrogen/Deuterium exchange for CD2AP SH3-A and C domains. In the SH3-C domain, all residues located in secondary structure elements, except for the loops, display ΔG_{ex} values similar to the global unfolding energy (ΔG_U) obtained by DSC. This result indicates that most regions in SH3-C, except for the flexible loops, exchange by a global unfolding reaction. In contrast, SH3-A displays two different types of residues with different ΔG_{ex} values. The first one forms the cooperative core of the protein and is constituted by the β_1 to β_4 strands with ΔG_{ex} values similar to the ΔG_U obtained by DSC. The second region constitutes the 3^{10} helix and the β_5 strand showing lower ΔG_{ex} values, indicating that H/D exchange in this region of the protein takes place by local or partial unfolding. Moreover, measurable protection is found for His14 in the RT loop and Val29 and Lys30 in the n-Src loop in SH3-A, indicating the presence of an extended hydrogen bonding network in both loops with lower flexibility, in agreement with the higher enthalpy for this domain as mentioned above.

The different intramolecular interaction networks in the SH3-A and C domains of CD2AP is also reflected in the different backbone flexibility exhibited by both domains. ^{15}N relaxation R2/R1 values report on low frequency motions of the protein backbone. Higher R2/R1 values and thus higher amplitude motions are observed in the RT, n-Src and distal loops of SH3-C, in agreement with the fewer stabilizing interactions present in these loops in this domain. SH3-A, in contrast, exhibits a lower flexibility in these loops. Furthermore, comparison of the R2/R1 values obtained for SH3-C at different pH values reveals similar backbone motions in the loops at low pH in SH3-A and C domains, in agreement with the similar thermal stability observed in these conditions. On the other hand, the SH3-A C-terminus, including the 3^{10} helix and the β_5 strand, displays higher amplitude motions in the low frequency timescale, an indicative of chemical exchange consistent with the lower ΔG_{ex} values obtained by H/D exchange.

We have thus established the link between structure, dynamics and thermodynamics. However, the biological repercussion of such different behaviour in solution of the neighbouring SH3 domains remains unclear. It was remarkable that the N-terminal domain of CD2AP has a higher stability than any other SH3 domain studied at physiological pH and 25°C, so a

comparison with other N-terminal SH3 domains was needed to check if this is a common characteristic of N-terminal SH3 domains. Nonetheless, such study has not been performed to date. Therefore, we used FOLDX to predict the thermal stability of the closely related SH3 domains of CIN85, a member of the same CD2AP family. Interestingly, we found a similar behaviour than in CD2AP, with the N-terminal SH3 domain being the most stable one, driven by a similar enthalpy value. This analysis suggests that the elevated stability for N-terminal SH3 domains might be a general characteristic of multiple domain containing proteins. On the other hand, CD2AP SH3-B and C displays similar ΔG that other SH3 domains [12, 13], indicating that SH3 domains located away from the N-terminus do not need such stabilization.

In chapters two and three we have described how subtle differences in the sequence and thus in the structure of the three SH3 domains of CD2AP result in a different stability and flexibility. Furthermore, we have established the molecular basis of such differences, and made the links with other SH3 containing proteins in order to relate their structure, dynamics and thermodynamics with the functions in which they are involved. Once we have achieved our first objective of the thesis, we directed our research towards the characterization of their interactions with natural targets. Ubiquitin is implicated in a wide variety of cellular processes. One of them is the internalization of cell surface proteins into the endocytic pathway, where the CIN85/CD2AP family is also involved. Recently, it was found that a subset of SH3 domains constitutes a new, distinct type of ubiquitin-binding domains [7]. The structural basis for the interaction between Sla1 SH3-3 and ubiquitin was established, pointing at Phe409 as the key residue in SH3 for ubiquitin binding [15]. Further studies have revealed that the three SH3 domains of CIN85 are also capable of ubiquitin binding [8], and mutation of the key Phe to tyrosine was found to abrogate ubiquitin binding. Since all three CD2AP SH3 domains posses a Phe residue at the equivalent position, we tested possible ubiquitin binding by performing fluorescence, ITC and NMR detected titrations. We found that the three domains bind to ubiquitin with different affinities. Moreover, while SH3-A and B display similar chemical shift perturbations (CSP) in the NMR titration experiments reflecting an alike orientation of both interacting molecules as in the Sla1 SH3-3:ubiquitin complex, SH3-C domain presents a different CSP profile of residues affected by ubiquitin binding. Particularly, the region around the supposed key phenylalanine at position 322 does not show significant changes in chemical shift upon ubiquitin, and different regions in ubiquitin are targeted as well. These observations were suggesting a possible alternative mode of ubiquitin binding by SH3 domains and thus triggered the determination of the high resolution structure of both type of complexes.

NMR titrations of SH3-A with ubiquitin and *visa-versa* show some residues displaying intermediate exchange, while SH3-C titrations show all residues in fast exchange indicating a substantial lower affinity for ubiquitin for the third SH3 domain. As a consequence, the measurement of intermolecular NOes is limited by the inherent insensitivity of the NOESY experiments and the low concentration of complex in solution. At the same time, extraction of orientational restraints from RDCs or CSA in weak complexes is compromised by several complications. This is essentially because under these conditions the complex cannot be isolated experimentally, and thus the measured RDCs report on a combination of both free and bound forms. In order to elucidate the high resolution structure of CD2AP SH3-C in complex with ubiquitin, we have developed a novel methodology that allowed us to address the experimental limitations described above and that can be generalized for the measurement, analysis and interpretation of RDCs for the refinement of the structure of weak macromolecular complexes. The method is based on three different aspects: (i) Differential isotopic labeling of SH3-C (^{13}C-^{15}N labelled) and ubiquitin (^{15}N labelled). This allowed the use of isotopically filtered experiments that permit the RDC measurements of both components in the same sample. (ii) RDC measurement for both free forms and different mixtures of both partners with known population of complex in the same sample. (iii) Global fitting of all the experimentally measured RDCs to determine the global scaling parameters for each mixture and the bound-form values for both partners from linear extrapolation. The bound-form RDCs of both partners can be in this way incorporated into the structure calculation protocol.

It is important to highlight the importance of using a titration approach for the determination of RDCs from weak complexes. As described before, the bound-form RDCs are extrapolated, ensuring that accurate alignment tensors are determined for the precise determination of weakly macromolecular complexes. In order to show the consequences of the use of RDCs determined in non-saturating conditions, we simulated three conditions of incomplete saturation: 90%, 80% and 70% saturation. The optimal orientation of the three complexes was determined and clearly shows the incorrectness of the structures calculated. It is therefore extremely risky to exploit RDC data without using the titration approach proposed here, even when the data is measured in the range of 70% to 100% and no consideration is taken about the actual complex and free populations. The importance of this observation should not be underestimated, as a structure calculated from RDC measured at 80% of saturation can include errors of 40-50° in the orientation of both partners, and these errors cannot be identified on the basis of data reproduction of the determined structure. We have therefore developed a robust approach, described in chapter four, capable of the structure determination of weak macromolecular complexes, and expect this method to contribute further to our understanding of the molecular basis of cellular processes.

The resulting high-resolution structure of CD2AP SH3-C in complex with ubiquitin obtained using this new methodology is importantly different from the Sla1 SH3-3:Ubiquitin [15] and the CIN85 SH3-C:ubiquitin structure [8], what asked for a deeper analysis of the ubiquitin binding by the SH3 domains of CD2AP, described in chapter 5. The CD2AP SH3-C:ubiquitin structure shows that, in agreement with the chemical shift perturbation data, Phe322 is placed at the edge of the binding interface. Mutation of Phe322 to a tyrosine in CD2AP SH3-C does not abolish ubiquitin binding, as observed by fluorescence, ITC and NMR. Moreover, the latter experiments reveal that the Phe322 to tyrosine CD2AP SH3-C mutant targets the same residues in ubiquitin than the WT protein. On the contrary, NMR titration experiments of CD2AP SH3-A showed similar residues affected than Sla1 SH3-3 and CIN85 SH3-C as it is described in [8]. We therefore also calculated the solution structure of the complex between SH3-A and ubiquitin using RDCs to confirm that this binding mode is similar to that adopted by Sla1 SH3-3. The structure indeed shows a similar orientation than the Sla1 SH3-3 complex, thus confirming the CSP data. Moreover, mutation of Phe53 (equivalent to Phe409 in Sla1 SH3-3) to a tyrosine abolishes ubiquitin binding. These results demonstrate that the two SH3 domains from the same adaptor protein show alternative modes of ubiquitin binding; while CD2AP SH3-A binds in a similar mode than Sla1 SH3-3, CD2AP SH3-C binds ubiquitin in a new distinct mode.

Both ubiquitin binding modes have been rationalized in order to establish the molecular basis that drives both different orientations. CD2AP SH3-A and C domains bind to the same face of ubiquitin, which involves the Ile44 patch, but differ mainly in the type of residues surrounding this area. While Sla1 SH3-3 and CD2AP SH3-A target the Leu8 and Val70 binding region, the binding interface for CD2AP SH3-C moves away from the Leu8 region more direction the C-terminal di-glycine patch in ubiquitin. The analysis of the surface of these domains reveals that CD2AP SH3-A and Sla1 SH3-3 present positive, negative and non-polar areas in the binding interface, while CD2AP SH3-C shows a mainly negatively charged surface. This results in different contacts established between polar residues in the RT and n-Src loops with the ubiquitin C-terminus.

The importance of the ubiquitin C-terminus is not novel in SH3:ubiquitin binding, as it was recently shown that Nck SH3-3 domain interacts with ubiquitin mainly targeting the ubiquitin C-terminus [16]. However, a different binding mode is expected in this case as different areas in the SH3 domain are affected upon ubiquitin binding. The Endophilin-A SH3 domain is not able to bind ubiquitin, but instead it binds to the Parkin ubiquitin like protein (Ubl) [17]. This complex also shows direct interactions with the Ubl C-terminus, but in this case the orientation of both partners is significantly different than that for Sla1 SH3-3 or CD2AP SH3-C.

Interestingly, differences in charged residues and in the length of the loops are found from comparison of the amino acid sequences of CD2AP SH3-A and C, Sla1 SH3-3 and CIN85 SH3-C domains. The CD2AP SH3-C domain displays a bigger negatively polar area in the RT loop that results in favourable interactions with Arg42 in ubiquitin that are not observed in CD2AP SH3-A and Sla1 SH3-3 complexes. This interaction causes a displacement of the interface of the SH3 domain towards the C-terminus of ubiquitin. Bigger differences are observed in the n-Src loop. CD2AP SH3-C contains a longer loop with a negatively polar surface allowing contacts with the Arg74 and the diglycine patch in the C-terminus of ubiquitin. CD2AP SH3-A and Sla1 SH3-3 display a shorter n-Src loop with positive and negatively polar residues on its surface that are not involved in contacts to the ubiquitin C-terminus. To rationalize the importance of the n-Src loop over the RT loop we constructed two SH3-C mutants; one in which Thr283, located in the RT loop, was mutated into an alanine, and one in which Glu302 was mutated into a lysine together with a deletion of Thr303, to mimic more or less the n-Src loop in Sla1 SH3-3. NMR detected titrations showed that the Thr283Ala mutation does not affect ubiquitin binding, similarly than the glutamate to alanine mutation in Sla1 SH3-3 [7]. In contrast, the Glu320Lys + Thr303 deletion mutant abolishes ubiquitin binding. The effect of this mutation was rationalized calculating a structural model with the program Modeller[17][16]. A shorter n-Src loop, together with the change in charge moves the loop further away from the ubiquitin C-terminus. At the same time this mutation eliminates the contact between Thr303 and Gly75 in ubiquitin and hinders the interaction between Gly304 and Arg74 due to electrostatic repulsions. This points out that the length and charge of the n-Src loop are the main determinants of the CD2AP SH3-C ubiquitin binding mode, and thus any SH3 domain displaying a similar length and charge in the n-Src loop might bind ubiquitin in an alike orientation. Nevertheless, the SH3:ubiquitin interaction is driven by tertiary contacts that depend on the overall charge and size of the whole binding interface.

Prior studies showed that CIN85 SH3-C domain bind ubiquitin in a similar way than Sla1 SH3-3 [8]. However, CIN85 SH3-C titrations with ^{15}N-labelled C-terminal His-tagged ubiquitin performed in this thesis revealed a very similar chemical shift perturbation pattern than that exhibited by CD2AP SH3-C. Interestingly, it has been described that CIN85 SH3-C presents two distinct conformations when bound to ubiquitin [18]. The first binding mode observed by CMPG relaxation dispersion overlaps significantly with the chemical shift perturbation pattern described above. Moreover, no contact involving the key Phe322 is observed in this orientation, suggesting that CIN85 SH3-C is able to bind C-terminal tagged ubiquitin in the same orientation than CD2AP SH3-C, what is not surprising due to their high sequence homology (71% of similarity and 50% of identity), especially in the n-Src loop, which is of the same length and

charge. This is in contradiction to the structure of the CIN85 SH3-C:Ubiquitin complex obtained by Forman-Kay and co-workers. In their work the F52Y, F151Y, F322Y triple mutant in a full length CIN85 construct did not eliminate CIN85 ubiquitination in *in vivo* ubiquitination experiments. Actually this observation is in agreement with our findings that like for CD2AP, also the mutation of Phe322 to tyrosine in CIN85 SH3-C does not lead to abolishing of ubiquitin binding.

Another key feature of the CD2AP SH3-C:ubiquitin binding mode is the importance of a C-terminal tag in ubiquitin for the binding affinity. Titrations of CD2AP SH3-C with C-terminal His-tagged ubiquitin results in an increase in affinity by one order of magnitude compared with non-tagged or N-terminal His-tagged ubiquitin. Such effect was previously reported for CIN85 SH3-C and amphiphysin SH3 domains [7], although in this case a GST-tag was present at the C-terminus of ubiquitin. These three SH3 domains all posses a longer and negatively charged n-Src loop, suggesting that they share the same ubiquitin binding mode. Moreover, the higher affinity to C-terminal tagged ubiquitin constitutes a feature of this binding mode, as it has not been observed for CD2AP SH3-A or Sla1 SH3-3. However, the biological role of this new ubiquitin binding mode is still unclear. While CD2AP SH3-A and B are able to tightly bind free ubiquitin, CD2AP SH3-C requires the ubiquitin molecule to be attached via its C-terminus to a protein to interact tightly. CD2AP, as well as CIN85, are ubiquitinated through interaction with the Cbl E3 ubiquitin protein ligase [19], constituting possible targets for ubiquitin binding by the CD2AP SH3-C domain.

We have shown in chapter five that the different ubiquitin binding capabilities of the SH3 domains of CD2AP could be related with the intrinsic flexibility of the endocytic pathway, were multiple low affinity interactions are needed to orchestrate the correct signal pathway inside the cell. The diversity in type of interactions in which CD2AP is involved and the fact that it plays a role in many cellular processes has revealed the necessity of a complete understanding of its behaviour inside the cell. The research described in this thesis aimed to shed light on the function of each SH3 domain inside CD2AP, as well as to establish the molecular basis that underlie their biological role inside the cell.

This thesis has shown the power of modern NMR in the field of (structural) biology and that this biophysical technique is an inevitable tool for unravelling the intricate interplay between structure, stability, dynamics and interaction of proteins.

References

1. Dustin, M.L., et al., *A novel adaptor protein orchestrates receptor patterning and cytoskeletal polarity in T-cell contacts.* Cell, 1998. **94**(5): p. 667-77.
2. Dikic, I., *CIN85/CMS family of adaptor molecules.* FEBS Lett, 2002. **529**(1): p. 110-5.
3. Moncalian, G., et al., *Atypical polyproline recognition by the CMS N-terminal Src homology 3 domain.* J Biol Chem, 2006. **281**(50): p. 38845-53.
4. Morita, E., et al., *Human ESCRT and ALIX proteins interact with proteins of the midbody and function in cytokinesis.* EMBO J, 2007. **26**(19): p. 4215-27.
5. Ababou, A., M. Pfuhl, and J.E. Ladbury, *Novel insights into the mechanisms of CIN85 SH3 domains binding to Cbl proteins: solution-based investigations and in vivo implications.* J Mol Biol, 2009. **387**(5): p. 1120-36.
6. Kirsch, K.H., et al., *The adapter type protein CMS/CD2AP binds to the proto-oncogenic protein c-Cbl through a tyrosine phosphorylation-regulated Src homology 3 domain interaction.* J Biol Chem, 2001. **276**(7): p. 4957-63.
7. Stamenova, S.D., et al., *Ubiquitin binds to and regulates a subset of SH3 domains.* Mol Cell, 2007. **25**(2): p. 273-84.
8. Bezsonova, I., et al., *Interactions between the three CIN85 SH3 domains and ubiquitin: implications for CIN85 ubiquitination.* Biochemistry, 2008. **47**(34): p. 8937-49.
9. Slupsky, C.M., et al., *Smartnotebook: a semi-automated approach to protein sequential NMR resonance assignments.* J Biomol NMR, 2003. **27**(4): p. 313-21.
10. Wuthrich, K., *Protein structure determination in solution by nuclear magnetic resonance spectroscopy.* Science, 1989. **243**(4887): p. 45-50.
11. Sadqi, M., et al., *pH dependence of the hydrogen exchange in the SH3 domain of alpha-spectrin.* FEBS Lett, 2002. **514**(2-3): p. 295-9.
12. Filimonov, V.V., et al., *A thermodynamic analysis of a family of small globular proteins: SH3 domains.* Biophys Chem, 1999. **77**(2-3): p. 195-208.
13. Knapp, S., et al., *Thermal unfolding of small proteins with SH3 domain folding pattern.* Proteins, 1998. **31**(3): p. 309-19.
14. Guerois, R., J.E. Nielsen, and L. Serrano, *Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations.* J Mol Biol, 2002. **320**(2): p. 369-87.
15. He, Y., L. Hicke, and I. Radhakrishnan, *Structural basis for ubiquitin recognition by SH3 domains.* J Mol Biol, 2007. **373**(1): p. 190-6.

16. Kang, J., et al., *Distinct interactions between ubiquitin and the SH3 domains involved in immune signaling.* Biochim Biophys Acta, 2008. **1784**(9): p. 1335-41.
17. Trempe, J.F., et al., *SH3 domains from a subset of BAR proteins define a Ubl-binding domain and implicate parkin in synaptic ubiquitination.* Mol Cell, 2009. **36**(6): p. 1034-47.
18. Korzhnev, D.M., et al., *Alternate binding modes for a ubiquitin-SH3 domain interaction studied by NMR spectroscopy.* J Mol Biol, 2009. **386**(2): p. 391-405.
19. Haglund, K., et al., *Cbl-directed monoubiquitination of CIN85 is involved in regulation of ligand-induced degradation of EGF receptors.* Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12191-6.

Chapter 7

Resumen y conclusiones/ Summary and conclusions

RESUMEN Y CONCLUSIONES

El trabajo presentado en esta memoria tenía como objetivo principal la caracterización de los dominios SH3 de la proteína CD2AP, así como de sus interacciones con dianas naturales, en especial ubiquitina, un ligando natural implicado en gran variedad de funciones dentro de la célula, tales como la regulación post-transcripcional de proteínas, señalización celular o transporte a través de la membrana. Este objetivo se consiguió a través de la determinación estructural de las formas libres de los tres dominios SH3 mediante resonancia magnética nuclear (RMN), y la caracterización dinámica y termodinámica de estos mediante RMN y calorimetría diferencial de barrido (CDB). El estudio de las interacciones con ubiquitina se llevó a cabo también mediante RMN tras desarrollar una metodología novedosa para la obtención de la estructura tridimensional de alta resolución de complejos de baja afinidad, como el formado por el tercer dominio SH3 de CD2AP y ubiquitina.

De forma más detallada el trabajo desarrollado se resume en:

1. Se clonaron el segundo y tercer dominio SH3 de CD2AP (SH3-B y C) y se purificaron los tres dominios SH3 de CD2AP en condiciones tanto de abundancia isotópica natural como enriquecidos isotópicamente en ^{13}C y/o ^{15}N.
2. Se llevaron a cabo estudios de estabilidad mediante CDB para determinar las condiciones más estables para cada uno de los dominios y para caracterizar su termodinámica de desplegamiento.
3. Se determinaron las estructuras tridimensionales de los tres dominios SH3 mediante la aplicación de distintas metodologías de RMN, tales como experimentos heteronucleares bi- y tridimensionales, acoplamientos dipolares residuales y experimentos homonucleares bidimensionales, lo que nos ha permitido interpretar las diferencias en estabilidad existentes entre ellos mediante una aplicación de cálculo computacional de las contribuciones a la estabilidad a partir de la estructura tridimensional. De esta forma se pudieron establecer también las diferencias estructurales que puedan afectar a las interacciones con sus dianas naturales. Los resultados obtenidos en la determinación de la estructura tridimensional del dominio SH3-C han sido recogidos en una publicación científica en la revista *Journal of Biomolecular NMR* (Ortega-Roldan *et al*. 2007).
4. Se llevó a cabo un estudio de la dinámica y termodinámica de la cadena principal mediante relajación de ^{15}N e intercambio hidrógeno deuterio para comprender la flexibilidad exhibida por estos dominios en disolución así como sus determinantes moleculares y poder relacionarla con la función biológica que desempeñan dentro de la célula. Los resultados descritos en los

apartados 1 a 3 han sido enviados para su publicación en la revista *Protein Science* (Ortega-Roldan *et al.* 2010).

5. Con el fin de poder abordar la determinación estructural de complejos macromoleculares de baja afinidad se desarrolló una metodología novedosa que emplea acoplamientos residuales dipolares medidos mediante espectroscopía de resonancia magnética nuclear. Este estudio se recoge en una publicación científica en la revista *Nucleic Acid Research* (Ortega-Roldan *et al*, 2009).
6. Por último, se caracterizaron las interacciones de los tres dominios SH3 de CD2AP con ubiquitina. Se elaboró un modelo estructural del complejo formado por el dominio SH3-A y ubiquitina y se determinó la estructura tridimensional de alta resolución del complejo formado por el dominio SH3-C y ubiquitina. Para comprender las diferencias entre ambas estructuras se elaboraron mutantes y se llevaron a cabo titulaciones monitorizadas mediante RMN, ITC y fluorescencia. Este análisis se completó con un estudio comparativo de todos los dominios SH3 capaces de interaccionar con ubiquitina. Los resultados obtenidos en este estudio han sido enviados para su publicación en la revista *The Journal of Biological Chemistry*.

Las conclusiones más importantes derivadas del trabajo presentado en esta memoria se enumeran a continuación:

1. La caracterización de la estabilidad de los tres dominios SH3 de CD2AP mediante CDB ha revelado una elevada estabilidad térmica del dominio SH3-A, asociada a su ubicación en el extremo N-terminal de CD2AP, mientras que los dominios SH3-B y C presentan una estabilidad semejante al resto de dominios SH3 estudiados hasta la fecha.
2. Las estructuras tridimensionales de alta resolución de los tres dominios revelan una gran homología estructural entre ellos, aunque se observan diferencias en la distribución de las cargas en su superficie, así como en la longitud del lazo n-Src.
3. El análisis de las estructuras mediante un algoritmo de predicción de las contribuciones a la energía de Gibbs (FoldX) demuestra que la mayor estabilidad del dominio SH3-A es producto del establecimiento de un mayor número de interacciones intramoleculares dentro de la estructura. Estudios de relajación de ^{15}N y de intercambio Hidrógeno/Deuterio han demostrado que estas interacciones conllevan asimismo una menor flexibilidad de los lazos RT, n-Src y distal.
4. Hemos desarrollado un método para la determinación de la estructura tridimensional de complejos débiles mediante la medida de acoplamientos dipolares residuales usando RMN.

5. Existe un modo de unión alternativo entre el dominio SH3-C de CD2AP con ubiquitina diferente al establecido para la interacción entre los dominios SH3-A y B de CD2AP y SH3-3 de Sla1 con ubiquitina
6. El dominio SH3-C de CD2AP presenta afinidad preferencial por ubiquitina cuyo extremo C-terminal se encuentra inmovilizado, lo cual constituye un comportamiento diferencial a los dominios SH3-A y B de CD2AP y SH3-3 de Sla1, lo que podría estar relacionado con su función biológica.

SUMMARY AND CONCLUSIONS

The research described in this thesis aimed the characterisation of the SH3 domains of CD2AP, as well as their interactions with their natural targets, specially ubiquitin, a natural target involved in a wide variety of cellular activities ranging from transcriptional regulation to cell signalling and membrane trafficking. This aim was achieved through the structural determination of the free forms of the three SH3 domains using Nuclear Magnetic Resonance spectroscopy (NMR), and the dynamic and thermodynamic characterisation of these domains by Differential Scanning Calorimetry (DSC) and NMR. Ubiquitin interactions with the SH3 domains of CD2AP have also been studied by NMR. The development of a novel approach to determine the high resolution structure of low affinity macromolecular complexes has allowed us to elucidate the structure of the complex between the third SH3 domain of CD2AP and ubiquitin.

A detailed summary of the work described in this thesis is given below:

1. The second and third SH3 domains of CD2AP (SH3-B and SH3-C) were subcloned, and the three SH3 domains of CD2AP were obtained in unlabelled as well as isotopically enriched in ^{13}C and/or ^{15}N, except for SH3-B, which due to its instability could not be obtained isotopically labelled.
2. Differential Scanning Calorimetry (DSC) was used in order to establish the most stable conditions for each domain, as well as for a thermodynamic characterisation of their unfolding processes.
3. The high resolution structures of the three SH3 domains were determined using different NMR methodologies, such as heteronuclear two- and three-dimensional experiments, residual dipolar couplings and homonucelar two-dimensional experiments. This study allowed us to rationalize their differences in stability by using a computational algorithm that calculates the energetic contributions to the stability from the three dimensional structure. Consequently, the structural differences affecting their interactions with natural targets were also established. The high resolution structure determination of SH3-C has been published in the *Journal of Biomolecular NMR* (Ortega-Roldan *et al*. 2007).
4. The dynamics and thermodynamics of the protein backbone were studied using NMR ^{15}N relaxation measurements and Hydrogen/Deuterium exchange in order to understand the flexibility exhibited in solution of domains SH3-A and C. The molecular determinants and biological relevance were also analysed. The results mentioned in points 1 to 3 have been sent for publication in *Protein Science* (Ortega-Roldan *et al*. 2010).
5. In order to address the structure determination of weak macromolecular complexes, a new methodology has been

developed exploiting the orientational information available from NMR residual dipolar couplings. This new approach has been published in the journal *Nucleic Acid Research* (Ortega-Roldan *et al*. 2009).

6. Finally, the interactions between the three SH3 domains of CD2AP and ubiquitin were studied. A data-based structural model of the complex between SH3-A and ubiquitin was determined, and the high resolution structure of the complex between SH3-C and ubiquitin was elucidated. The different binding modes found were further studied by mutagenesis and NMR detected titrations. This analysis was completed with a comparative study of all ubiquitin interacting SH3 domains. These results have been sent for publication in *The Journal of Biological Chemistry*.

The most relevant conclusions from the work described in this thesis are enumerated below:

1. The characterisation of the stability of the three SH3 domains of CD2AP using DSC has revealed a higher thermal stability of the N-terminal SH3 domain of CD2AP, while SH3-B and C domains display a similar stability than other SH3 domains described in literature up to date.
2. The high resolution structures of the three domains show a high structural homology, although key differences are found in the charge distribution and in the length of the n-Src loop.
3. Analysis of the structures using a computational algorithm for prediction of the contributions to the Gibbs energy (FoldX) shows that the higher stability of SH3-A is driven by a higher number of non-bonded intramolecular interactions. ^{15}N relaxation and Hydrogen/Deuterium exchange studies display that these interactions are also responsible of the lower flexibility of the RT, n-Src and distal loops in SH3-A.
4. We have developed a new methodology for the structure determination of weak macromolecular complexes using NMR residual dipolar couplings.
5. CD2AP SH3-C domain adopts a new ubiquitin binding mode, different from that exhibited by the CD2AP SH3-A and B domains, as well as Sla1 SH3-3.
6. The SH3-C domain of CD2AP presents a preferential affinity for ubiquitin with an elongated or immobilised C-terminus. This effect is not observed for the CD2AP SH3-A and Sla1 SH3-3 domains, and therefore likely to be related to a different biological role.

ABBREVIATIONS

1D	Mono-dimensional
2D	Two-dimensional
3D	Three-dimensional
AIR	Ambiguous interaction restraint
$\mathbf{A_r}$	Rombic component of the alignment tensor
$\mathbf{A_a}$	Axial component of the alignment tensor
BEST	Band-selective excitation short-transient
BMRB	Biological Magnetic Resonance Bank
CB	Column buffer
CD2AP	CD2 associated protein
CIN85	Cbl-interacting protein of 85 kDa
CMS	Cas ligand with multiple SH3 domains
CPMG	Carr-Purcell-Meiboom-Gill based T_2 relaxation dispersion
CS	Chemical shift
CSP	Chemical shift perturbation
$\mathbf{C_p}$	Heat capacity
DLS	Dynamic Light Scattering
DSC	Differential Scanning Calorimetry
DTT	1,4-Dithiothreitol
$\mathbf{D_{ij}^{exp}}$	Experimental residual dipolar coupling
$\mathbf{D_{ij}^{back}}$	Backcalculated residual dipolar coupling
$\mathbf{D_{ij}^{free}}$	Residual dipolar coupling of the free form of the protein
$\mathbf{D_{ij}^{bound}}$	Residual dipolar coupling of the bound form of the protein

$^{1}D_{NH}$	N amide - H amide coupling
$^{1}D_{C\alpha C'}$	Cα - C carbonilic coupling
$^{2}D_{HN\ C'}$	H amide - C carbonilic coupling
$^{1}D_{C\alpha H\alpha}$	Cα - Hα coupling
EFG	Epidermal growth factor
EX2	Type 2 exchange mechanism
HSQC	Heteronuclear single quantum correlation
IPTG	Isopropyl-B-D-Thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
K_b	Thermodynamic binding constant
k_{ex}	Exchange rate obtained from H/D exchange
K_d	Thermodynamic dissociation constant
MD	Molecular dynamics
NMR	Nuclear Magnetic Resonance
NOe	Nuclear Overhauser effect
NOESY	Nuclear Overhauser enhancement spectroscopy
PDB	Protein Data Bank
PFG	Pulse Field Gradient
pI	Isoelectric point
p_{bound}	Population of the bound form of the protein
RDC	Residual dipolar coupling
RMSD	Root mean square deviation
RTK	Receptor tyrosine kinase
S	Order parameter obtained from Model Free analysis
SH3	Src-Homology domain 3

T	Temperature
TOCSY	Total correlation spectroscopy
T_m	Transition temperature between the native and unfolded states
UBD	Ubiquitin binding domain
Ubl	Ubiquitin like domain
WT	Wild type
ΔG	Gibbs energy change of unfolding
ΔG_{ex}	Gibbs energy change obtained from H/D exchange
ΔH_b	Enthalpy change of binding
ΔH	Enthalpy change of unfolding
ΔC_p	Heat capacity difference between the native and unfolded states