

NMR-based docking of protein-protein complexes

The human UbcH5B-CNOT4 ubiquitination complex

Dokking van eiwit-eiwit complexen met behulp van NMR data

Het humane UbcH5B-CNOT4 ubiquitineringscomplex

(met een samenvatting in het Nederlands)

Proefschrift ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de Rector Magnificus, Prof. Dr. W.H. Gispen, ingevolge het besluit
van het College voor Promoties in het openbaar te verdedigen op
maandag 14 juni 2004 des middags te 2.30 uur.

door

Cyril Dominguez

Geboren op 6 augustus 1975,

te Marseille, Frankrijk

Promotoren: Prof. Dr. R. Boelens
Prof. Dr. R. Kaptein

Co-promotor: Dr. A. M. J. J. Bonvin

Allen zijn verbonden aan het Bijvoet Centrum voor Biomoleculair Onderzoek,
Faculteit Scheikunde van de Universiteit Utrecht.

This work was supported by the Center for Biomedical Research (CBG) and the
Netherlands Organization for scientific research (NWO)

CONTENTS

General introduction	5
Chapter 1: Biological Introduction: RING domain proteins	9
Chapter 2: Methodological Introduction: Modeling of macromolecular interactions based on NMR data	33
Chapter 3: HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical data	51
Chapter 4: Solution NMR Structure of the E2 ubiquitin conjugating enzyme UbcH5B	67
Chapter 5: Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis and docking approaches	89
Appendix	113
Summary	117
Samenvatting	119
Résumé	121
Acknowledgments	123
Curriculum Vitae	125
List of publications	127

Beoordelingscommissie:

Prof. Dr. H. Th. M. Timmers
Department of Physiological Chemistry
University Medical Center Utrecht

Prof. Dr. A. J. R. Heck
Department of Biomolecular Mass Spectrometry
University of Utrecht

Prof. Dr. J. P. Kamerling
Department of Bio-organic Chemistry, Section of Glycoscience and Biocatalysis
University of Utrecht

Dr. B. Gilquin
Laboratory of Protein Structure
Centre d'Etude Atomique (CEA) Saclay

Paranimfen:

Klaartje Houben
Cedric Bernard

Thesis layout: AV-dienst Faculteit Scheikunde, Universiteit Utrecht

Cover Design: Eva van Dongen

Reproduction: PrintPartners Ipskamp

ISBN: 90-393-3721-7

General Introduction

The Human Genome Project has provided the sequence of roughly 30 000 genes but did not provide direct information on their function. To carry out its function, a protein generally interacts with other molecules and therefore an understanding of function depends on our ability to map the interaction networks that exist in different cellular processes. Interactions are critical for the formation of complexes and the viability of a cell. Furthermore, regulatory mechanisms, which control the interactions between proteins, are often achieved through interactions. Assuming that there are on average 5 interactions per protein, one can expect a minimum of 150 000 possible interactions in human cells (Figeys, 2003). Understanding the molecular and functional interactions among macromolecular complexes, as well as their changes associated with time, cell type or disease state will be invaluable for human health, and will have direct implications, for example, in pharmaceutical research to identify and select potential targets, and design efficient and specific drugs. Large-scale analysis and characterization of protein-protein interactions, i.e. the “interactome”, extends greatly our knowledge on macromolecular complexes and therefore on the proteome organization (Gavin and Superti-Furga, 2003). There are many ways to map and study macromolecular interactions that can vary from large-scale approaches at a cellular level to structural approaches at an atomic level. Large-scale approaches generally rely on two methods: molecular biology through yeast two-hybrid screening or tandem affinity purification to map interactions and mass spectrometry to identify the proteins involved (Figeys, 2003; Sobott and Robinson, 2002). A recent example of such a large-scale protein interaction mapping is the analysis of protein interactions in *Drosophila Melanogaster* (Giot et al., 2003) using yeast two-hybrid screening and probability statistics. This mapping identified with high confidence 4780 different interactions involving 4679 proteins. Biochemical and structural approaches, however, are only suitable to screen a small number of interactions, being part of a complete pathway, but they clearly give more insight into the molecular basis of these interactions. Mappings and structural studies of macromolecular interactions can therefore be performed in a complementary way: results of large-scale interaction mapping often provide a starting point for detailed biochemical and/or structural studies to gain more insight into interactions that are of particular relevance for medical and pharmaceutical research. Not all molecular interactions, however, can be approached by experimental structural studies of macromolecular complexes. Limitations of these methods are evident, especially in the case of weak and transient complexes. New and complementary methodologies have therefore been developed that can partially overcome this problem. In particular, docking and molecular dynamic simulations, performed *in silico*, can overcome experimental limitations and form currently an active area of research with good future prospects.

Scope of this thesis

This thesis describes structural and methodological aspects of the study of macromolecular complexes with an emphasis on NMR spectroscopy and focus on a specific and well-controlled biological pathway, the ubiquitination pathway. The general ubiquitination mechanisms are conserved over all eukaryotic organisms. Deregulation of the interactions involved in this pathway often leads to severe diseases among which cancer and various neurodegenerative diseases. The ubiquitination of a protein consists of a cascade of enzymatic reactions that results in the covalent attachment of the protein ubiquitin to a target protein. Attachment of multiple ubiquitin moieties to a protein (poly-ubiquitination) generally targets the protein for degradation by the 26S proteasome, while attachment of one ubiquitin (mono-ubiquitination) generally acts as a regulation event, in the same manner as phosphorylation or acetylation. The cascade involves a number of enzymes, and the various interactions between them and with the final substrate have been shown to be highly specific. The molecular basis of this specificity is, however, poorly understood. This thesis describes the development of the computational methods to study biomolecular interactions and their applications for the structure determination of a complex between the RING domain of a human transcription factor, CNOT4, and the ubiquitin conjugating enzyme (E2) UbcH5B.

In Chapter 1, a biological introduction is given that describes the RING domain family and its role in the ubiquitination pathway. In this Chapter, structural, biological, chemical, and medical aspects of this domain and important protein-protein interactions that involve the RING domain are discussed.

Chapter 2 describes various computational and experimental methods for studying macromolecular complexes at molecular level with a focus on nuclear magnetic resonance (NMR) and docking approaches. Limitations of these two approaches are underlined and new methodologies that combine them are discussed in details.

In Chapter 3, a new computational method for docking of biomolecules is presented that includes biochemical and/or biophysical information to drive the docking process. The method is successfully tested on several protein-protein complexes with known structure and for which either NMR or mutational interaction data was available.

Chapter 4 describes the solution structure of the E2 enzyme UbcH5B obtained from model-guided automated NOE assignment in combination with diffusion anisotropy derived from NMR relaxation measurements. The solution structure is compared with other E2 enzyme structures and differences are highlighted.

Finally, in Chapter 5, a structural model of the UbcH5B/CNOT4 complex involved in the ubiquitination pathway is described. By combining chemical shift perturbation mapping data and mutagenesis experiments, a precise definition of the binding interface of both partners is obtained. This information is used to dock the two proteins together using the docking approach described in Chapter 3. The resulting structural model is compared to the crystal structure of an homologous complex, UbcH7/c-Cbl (Zheng et al., 2000) in order to gain insight into the specificity of E2-E3 interactions.

References

- Figeys, D. (2003). Novel approaches to map protein interactions. *Curr Opin Biotechnol* *14*, 119-125.
- Gavin, A. C., and Superti-Furga, G. (2003). Protein complexes and proteome organization from yeast to man. *Curr Opin Chem Biol* *7*, 21-27.
- Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E., *et al.* (2003). A protein interaction map of *Drosophila melanogaster*. *Science* *302*, 1727-1736.
- Sobott, F., and Robinson, C. V. (2002). Protein complexes gain momentum. *Curr Opin Struct Biol* *12*, 729-734.
- Zheng, N., Wang, P., Jeffrey, P. D., and Pavletich, N. P. (2000). Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* *102*, 533-539.



Biological Introduction: RING domain proteins

Cyril Dominguez, Gert E. Folkers, and Rolf Boelens

Contribution to Handbook of Metalloproteins, volume 3, 338-351,
Wiley and Sons (2004)

Reproduced with permission of John Wiley and Sons, Ltd

Abstract

The RING domain is a cysteine-rich sequence motif that can bind two zinc atoms. In the canonical RING motif, also called the C3HC4 motif, one zinc is bound to four cysteines, and the other ion to three cysteines and a histidine. The tetrahedral coordination is atypical and referred to as a “cross-brace” motif. There are now more than 380 RING motifs identified in the human genome. The RING domain has been shown to mediate a crucial step in the ubiquitination pathway that targets protein substrates for degradation by the 26S proteasome. This pathway involves three enzymes and the RING finger proteins have been classified in this scheme as ubiquitin ligases. There are actually nine structures of RING domains solved by X-ray and NMR. The $\beta\alpha\beta$ fold of the RING domain is conserved among all structures. The structures of RING domains in complex with other proteins involved in the ubiquitination pathway provide considerable insight into the molecular basis of ubiquitination. Recently, ubiquitination has been shown to be not only a simple protein removal system but also an indispensable regulatory process. This is underscored by the observation that many diseases, like cancer and Parkinson’s disease, are due to mutations in RING domains that prevent an efficient ubiquitination and degradation process. In this review, we compare the different structures of RING domains, analyze their functional aspects through structural informations of RING domains in complex with other proteins and describe medical aspects of RING finger proteins involved in cancer and Parkinson’s disease.

Functional class

Eukaryotic zinc binding motif; this motif is related to the Zinc finger motif present in many transcription factors (Folkers et al., 2001; Schwabe and Klug, 1994), and was found initially in the *RING1* (Really Interesting New Gene) gene and therefore named RING finger (Lovering et al., 1993). Subsequently, this motif was detected in a number of unrelated proteins (Freemont et al., 1991; Lovering et al., 1993).

Initially the function of RING domains was not clear, although they were known to mediate protein-protein interactions and to be involved in a range of cellular processes, including development, oncogenesis, apoptosis, and viral replication (Borden, 2000; Borden and Freemont, 1996). By 1999, the function of the RING domain was clarified, with the observation that the RING domain of c-Cbl mediates a protein-protein interaction with proteins known to be involved in the protein ubiquitination and 26S proteasome degradation pathways (Joazeiro et al., 1999; Waterman et al., 1999; Yokouchi et al., 1999). Thereafter a similar function was deduced for a number of RING proteins (Lorick et al., 1999).

Occurrence

RING domains of RING finger proteins are one of the most common zinc binding motifs in eukaryotes (Saurin et al., 1996). There have been more than 380 RING motifs identified in the human genome and the number of sequences in all eukaryotes corresponding to a RING motif is currently more than 1200 (<http://www.sanger.ac.uk/Software/Pfam>) (Bateman et al., 2002). RING finger containing proteins can be found in a large variety of different species ranging from yeast to human including double strand DNA viruses and in all kind of cells or tissues (Freemont, 2000). RING proteins are not found in bacteria, which relates to their unique role in the ubiquitination pathway that is absent in prokaryotes.

Biological function

RING domains have been shown to mediate a crucial step in protein degradation. Figure 1 shows how a ubiquitin chain can be covalently linked to protein substrates (Borden, 2000; Freemont, 2000; Jackson et al., 2000; Joazeiro and Weissman, 2000). These poly-ubiquitinated substrates will then be recognized by the 26S proteasome and degraded by proteolysis (Glickman and Ciechanover, 2002; Pickart, 2001; Weissman, 2001). Three enzymes are involved in the ubiquitination pathway. Firstly, an E1 or ubiquitin activating enzyme forms a thiol ester with the carboxyl terminal group of the small protein ubiquitin at position Gly76. The ubiquitin is then transferred to the ubiquitin conjugating enzyme (E2). Finally, ubiquitin ligase (E3) transfers the ubiquitin from E2 to the target protein promoting the ubiquitination of the substrate. At least three different classes of E3 ligases have been found that mediate substrate ubiquitination. These E3 enzymes differ in the domain that recognizes the E2 enzymes, which can be a RING, a PHD (plant homeodomain) or a HECT (Homologous to E6AP COOH terminus) domain (Coscoy and Ganem, 2003; Glickman and Ciechanover, 2002; Pickart, 2001). In case of the HECT type E3 ligases, the small ubiquitin protein is captured from the E2, and covalently bound to the E3, before it is transferred to the substrate (Scheffner et al., 1995). In contrast, in the case of RING or PHD containing protein ligases, no evidence for a stable E3-ubiquitin intermediate exists; therefore, the E3 ligases that contain a RING or a PHD domain are thought to have only a role in bringing the E2 and the substrate together. In all cases, the interaction between the E2 and E3 enzymes is mainly accomplished by the HECT, PHD or RING domains.

It is now clear that protein degradation is an important feature in the regulation of cellular processes. Many vital functions like signaling, growth, transcription and DNA repair are all regulated by the 26S proteasome (Glickman and Ciechanover, 2002). Aberrant expression caused by deregulated protein degradation leads to severe diseases. Frequently RING finger proteins have been related to such diseases (see below) (Fang et al., 2003; Michael and Oren, 2003; Shtiegman and Yarden, 2003).

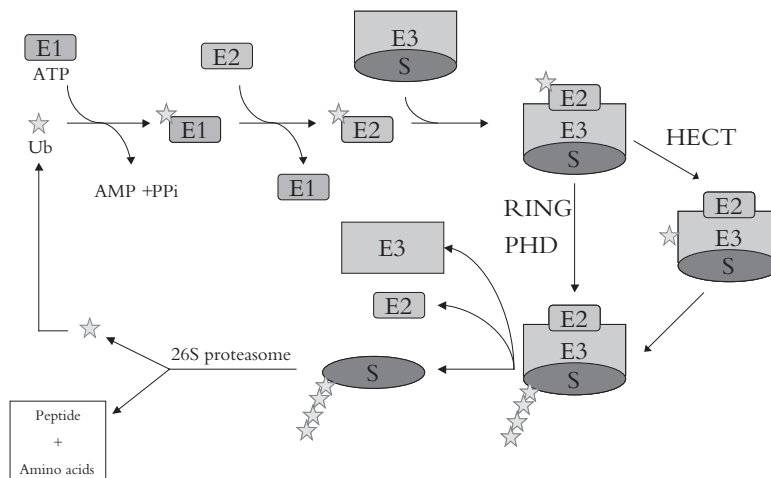


Figure 1: The ubiquitination pathway. Ubiquitin molecules are represented as stars, ubiquitin activating enzyme as E1, ubiquitin conjugating enzyme as E2, ubiquitin ligase as E3 and substrate as S.

Recently, another role for ubiquitination has been suggested. Whereas poly-ubiquitination is generally the marker on a substrate that will be degraded by the proteasome, a mono-ubiquitination of a protein can have a regulatory function as for example in signal transduction, transcription regulation, chromatin remodeling, and DNA repair (Deng et al., 2000; Kaiser et al., 2000; Kuras et al., 2002; Mallery et al., 2002; Wang et al., 2001). This new role suggests that the ubiquitination pathway is not only involved in degradation but might also regulate protein function. This protein modification can, like phosphorylation, acetylation and methylation, regulate protein activity, where the ubiquitin moiety is acting as a molecular switch in many important cellular processes.

Amino acid sequence information and topology

The only common feature of RING finger proteins is the presence of the cysteine-rich sequence motif that has similarities with Zinc fingers. This consensus RING motif, also called the C3HC4 motif, can be defined as a unique pattern of cysteine and histidine residues at defined positions in a peptide sequence, which is

Cys - X₂ - Cys - X₉₋₃₉ - Cys - X₁₋₃ - His - X₂₋₃ - Cys - X₂ - Cys - X₄₋₄₈ - Cys - X₂ - Cys, and where X can be any amino acid. It was suggested that the eight cysteine and histidine residues form a binding site for two zinc atoms. Since these transition metals can also be coordinated by other residues, it can be expected that variations on the RING consensus sequence are possible. Correspondingly, a second large group of RING finger proteins, the RING-H2

family, has been identified where the cysteine at position 4 is replaced by a histidine (Borden and Freemont, 1996). Also, a third consensus sequence for RING domains was found which consists of the C4C4 motif, in which all the zinc ligating residues correspond to cysteines (Hanzawa et al., 2001). In all three cases, the size of RING domains is approximately 70 amino acids (Figure 2) and representation of all three families have been shown to fold with a characteristic topology around the zinc ions. From sequence comparisons, it was proposed that the RING domain family also includes the U-box, a domain that is also involved in the ubiquitination pathway (VanDemark and Hill, 2002). However, this domain does not bind zinc atoms, but models predict a similar structure (Aravind and Koonin, 2000), although structure determination is required to confirm that this domain indeed adopts a similar fold.

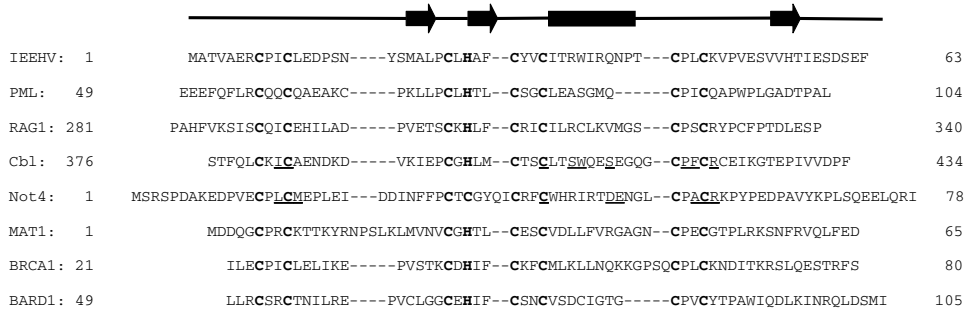


Figure 2: Sequence alignment of RING domains. Zinc ligands are displayed bold. Residues involved in the binding to an E2 enzyme (c-Cbl and CNOT4) are underlined. Secondary structure elements are based on the structure of IEEHV.

Activity test

The involvement of RING domains in the ubiquitination pathway is generally tested using an *in vitro* assay (Belz et al., 2002). The most common method is the so-called in-solution ubiquitination assay (Joazeiro et al., 1999; Koegl et al., 1999; Swanson et al., 2001), that uses ubiquitin, the E1, E2 and E3 enzymes, ATP and an ATP-regenerating system in a reaction mixture. In cases where the protein substrate for the ubiquitination is known, it is also added to the reaction. Ubiquitination will lead to an accumulation of high molecular weight polyubiquitin adducts bound to the substrate, or in the case of a substrate-independent assay, autoubiquitination of the E3 ligase (Albert et al., 2002; Joazeiro et al., 1999; Lorick et al., 1999). The characteristic ladder corresponding to multiple ubiquitin adducts is usually detected on polyacrylamide gel using western blot analysis (with anti-ubiquitin or anti-substrate antibodies) or autoradiography (with radiolabeled ubiquitin).

Protein production, purification, and molecular characterization

RING domains are frequently expressed and purified as Glutathione S transferase (GST)-RING fusion proteins (Gervais et al., 2001; Zheng et al., 2002; Zheng et al., 2000). However, also protocols using fusion constructs with the Maltose Binding Protein (MBP) or a His₆-Tag have been described (Bellon et al., 1997; Hanzawa et al., 2001). The recombinant RING proteins are overexpressed in *Escherichia coli*. Since zinc is needed for a proper folding, ZnCl₂ is added to the bacteria during the protein production. The GST-RING fusion proteins are purified using Glutathione affinity chromatography, whereafter the GST part is cleaved off by proteases. The RING domains or proteins are then further purified by standard methodology such as ion exchange and/or gel filtration chromatography. Buffers used for the purification should contain ZnCl₂, preferably a low concentration of phosphate, EDTA should be absent, and the pH should not be lower than 6–6.5 in order to prevent the protonation of the zinc coordinating residues and the subsequent loss of the metals, which generally leads to unfolding and precipitation of the RING domain.

Metal content

The RING domains comprise eight metal-binding ligands and bind two zinc(II) ions with each metal ion in a tetrahedral coordination. In a C3HC4 RING domain, one zinc ion is bound to four cysteines, and the other to three cysteines and a histidine, whereas in a RING-H2 domain both zinc ions are coordinated to three cysteines and a histidine. The metal content of RING domain proteins has been deduced by optical spectroscopy (Lovering et al., 1993) and by atomic absorption spectroscopy (Everett et al., 1993; von Arnim and Deng, 1993).

Zinc site geometries

The arrangement of the coordinating residues around the zinc atoms in the RING domain has often been erroneously ascribed to two tandemly arranged Zinc finger domains of the CCCH and CCCC type. However, the zinc ligation topology of RING fingers is quite distinct and is referred to as a “cross-brace” motif (Figure 3). In this motif, the first pair of ligands (Cys1 and Cys2) together with the third pair (Cys5 and Cys6) coordinates one zinc atom, and the second (Cys3 and His4) and fourth (Cys7 and Cys8) pairs bind the second zinc atom. In contrast, in the classical Zinc finger proteins, the first four coordinating residues would bind to one zinc atom and the next four to the next zinc atom. Sometimes, these sequential Zinc fingers form independent stable folds connected by flexible peptide segments as found in several transcription factors. In other cases a stable fold requires the close packing of both fingers (Folkers et al., 2001), as is the case in the DNA binding domain

of the steroid hormone receptor and the LIM domain (Dawid et al., 1998) (Figure 3). However, in the case of the “cross-brace” RING motif, it is clear that both zinc atoms are required for an intact structure.

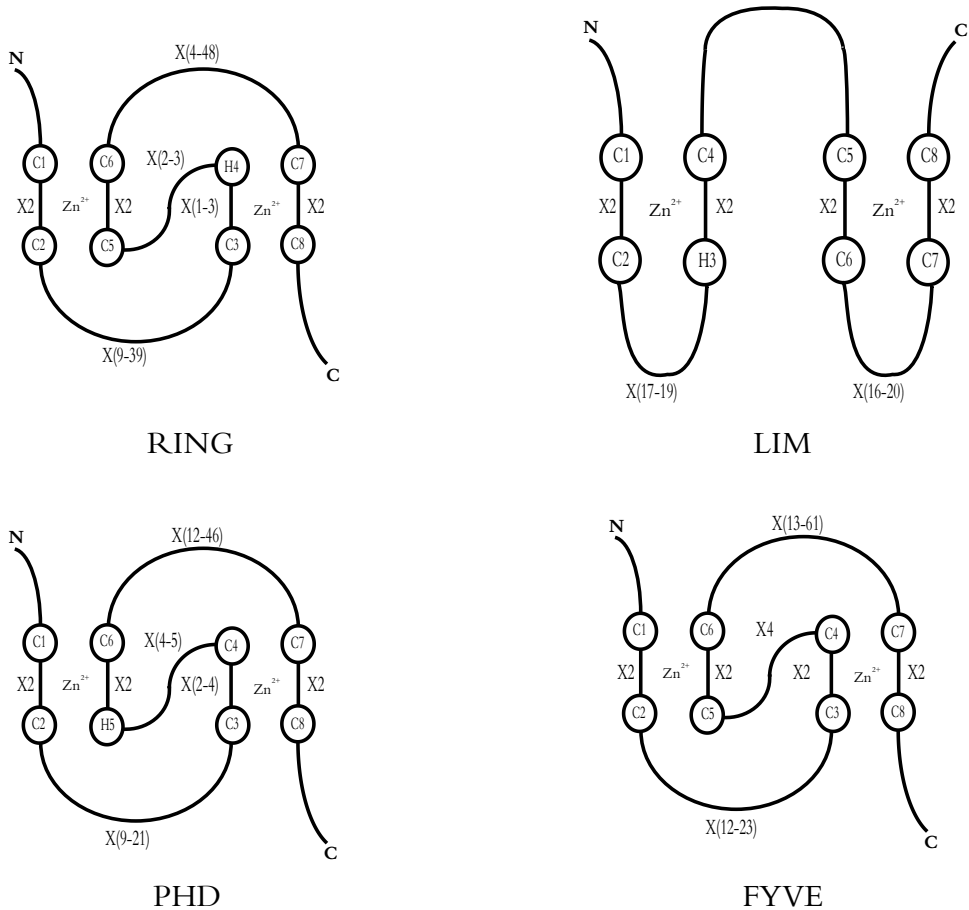


Figure 3: Zinc coordination topology for RING, LIM, PHD and FYVE domains.

This “cross-brace” motif is not unique for RING fingers, but can also be found in the plant homeodomain (PHD) (Aasland et al., 1995) and in the FYVE (Fab1p, YOTB, Vac1p, EEA1) domain (Misra and Hurley, 1999; Stenmark et al., 2002), a domain that was originally observed in the Fab1p (formation of aploid and binucleate cells), YOTB (Hypothetical *Caenorhabditis elegans* protein ZK632.12 in chromosome III.), Vac1p (also known as VSP19 involved in vacuolar segregation) and EEA1 (early endosome antigen 1) proteins (Figure 3).

The zinc ions of the RING domains are coordinated to the sulfur of the cysteines. The coordination to the histidine imidazole ring is unusual. Commonly in zinc fingers, the zinc coordinates to the NE2 of the imidazole group, but in the RING domains, coordination to the ND1 atom was demonstrated (Barlow et al., 1994; Bellon et al., 1997). This was explained by the close spacing between the ligands 3 and 4, which is conserved among all C3HC4 RING finger proteins, whereas, in most Zinc finger proteins, a more relaxed two-residue spacing is found. In the RAG1 (recombination-activating gene protein) structure (Bellon et al., 1997), the distances between the cysteine SG and the zinc atoms are between 2.23 and 2.36 Å, while the distance between the histidine ND1 and the zinc atom is 2.07 Å (Figure 4).

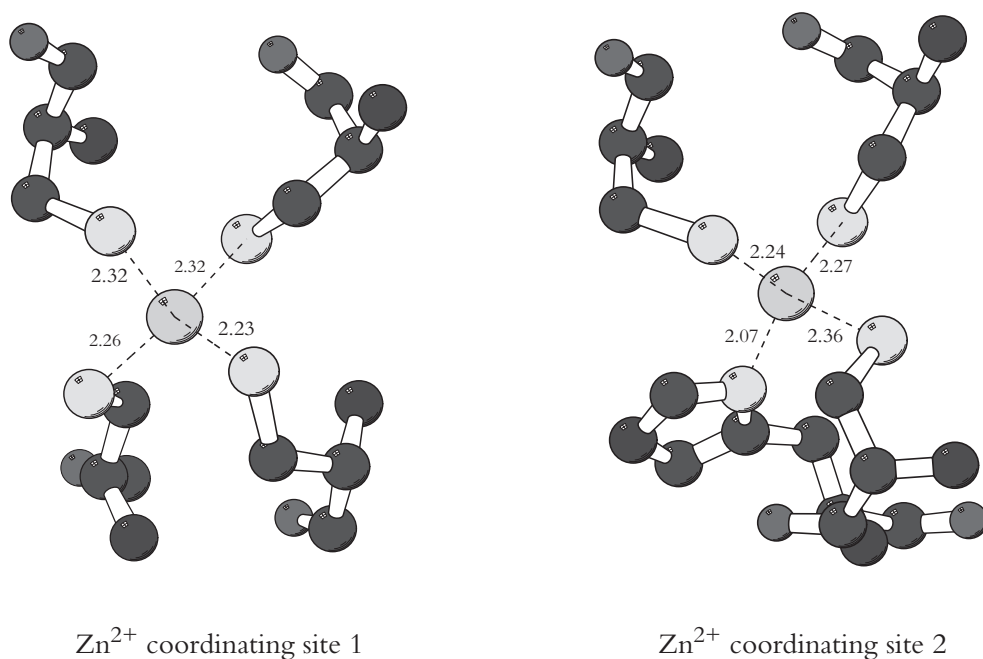


Figure 4: Zinc coordinating site geometry for the C4 site (site I) (left panel) and the C3H site (site II) (right panel) of the RAG1 RING domain. Zinc atoms, cysteine SG and histidine ND1 atoms are represented in grey. Distances in Å between the cysteine SG or histidine ND1 and the zinc ion are displayed. The figures have been generated with the program Molscript (Kraulis, 1991).

Spectroscopic and metal binding properties of RING finger domains

There are only a few spectroscopic techniques that enable the direct observation of zinc ions. In principle, zinc-EXAFS (Extended X-ray Absorption Fine Structure Spectroscopy) allows

the direct study of the metal binding properties of Zinc proteins (Garner et al., 1982) but this has not been applied thus far for RING fingers. The spectroscopic properties of RING domains can be investigated indirectly using a variety of spectroscopic techniques, due to the fact that one can substitute the zinc ion by other metals such as cobalt(II) or cadmium(II) (Lovering et al., 1993). The replacement of zinc by cobalt allows the observation of the metal site by UV/VIS and fluorescence spectroscopy. Two studies showed that the C3HC4 RING domains of BRCA1 (Breast cancer type 1) and hdm2 (human double minute 2 protein) bind the two cobalt atoms in a sequential manner, where the C4 site (site I) has a higher affinity than the C3H site (site II), which has been ascribed to an intrinsic differential stability of the two sites for the metal binding (Lai et al., 1998; Roehm and Berg, 1997). An alternate explanation, however, could be that this differential stability is directly related to a lower affinity of the histidine imidazole group for the cobalt ion, decreasing the affinity of the second site with respect to that of the first. Similarly, zinc can be replaced by cadmium and this allows the analysis of the coordination site by NMR. Hanzawa et al (unpublished data) studied the zinc-cadmium exchange by NMR titration experiments using the RING domain of CNOT4 (negative on TATA), which has an unusual C4C4 motif as confirmed by ^{113}Cd - ^1H HSQC experiments. In this case, it was found that the first site exchanges the zinc before the second site. These experiments demonstrate that the lower stability of the second metal binding site of the RING domain is not a general property of all RING domains.

X-ray and NMR structures of RING finger domains

First structural information about RING domains came from the analysis of the NMR chemical shifts of the RING domain of the immediate early EHV-1 (IEEHV-1) protein from equine herpes virus (Everett et al., 1993). Shortly thereafter, the three-dimensional structure of IEEHV was solved by the same group (Barlow et al., 1994) (Table 1). Since then, a number of structures of other RING domains have been solved by NMR and X-ray crystallography, both in the free state (Bellon et al., 1997; Borden et al., 1995; Gervais et al., 2001; Hanzawa et al., 2001) or in complexes with other proteins (Brzovic et al., 2001; Zheng et al., 2002; Zheng et al., 2000) (Table 1).

Overall description of the structure

Comparison of the nine RING domain structures shows that all RING fingers adopt a similar fold, but significant differences are present. All RING fingers display the so-called “cross-brace” motif for the zinc ligation. The overall structure is characterized by a $\beta\alpha\beta$ fold but the number of β -strands and the exact position of the β -sheet differ among the structures when analyzed with a DSSP algorithm (Kabsch and Sander, 1983) (Figure 5).

A common feature of all C3HC4 structures is the 14 Å distance between the two zinc atoms of the RING motif. In all structures, a hydrophobic cluster is found, that stabilizes the ternary structure. The main differences observed among the nine structures relate to the

presence and length of the secondary structure elements.

Structures of the free RING domains

The NMR structure of the IEEHV (Table 1) has a cross-brace coordination and a secondary structure that is composed of 3 β -strands and a central α -helix (Figure 5). The first loop (residues 1 to 19) contains the first pair of zinc ligating residues (Cys8 and Cys11). This loop is followed by two small β -strands (residues 19 to 21 and 26 to 28) connected by a short turn containing the second pair of zinc ligating residues (Cys24 and His26). The second β -strand is followed by an α -helix (residues 32 to 40) with the third pair of zinc ligating residues (Cys29 and Cys32) between both secondary structure elements. The C-terminal part of the structure consists of a long loop (residues 33 to 63) containing the fourth pair of zinc ligating residues (Cys43 and Cys46) and a third short β -strand (residues 53 to 55).

The C₃H₄C₄RING domain of the promyelocytic leukaemia proto-oncoprotein, PML, was the second RING structure that was determined by NMR (Borden et al., 1995) (Table 1). Although the metal ligation is similar to IEEHV and the distance between the two zinc atoms is conserved, the ternary structure is quite different from that of IEEHV, and the protein does not possess the central α -helix.

Table 1: Ring finger protein structures deposited in the PDB. Represented are the PDB code, the name of the protein, the type of RING, the technique used (X-Ray or NMR), the resolution of the structure (X-Ray) in Å or the backbone RMSD from mean structure (NMR) and the function of each RING.

PDB code	Protein name	Type of RING	Technique	Res / RMSD	Function
1CHC	IEEHV	C3HC4	NMR	0.55	Regulation of the equine herpes virus gene expression
1BOR	PML	C3HC4	NMR	0.88	Cellular defense mechanism
1RMD	RAG1	C3HC4	X-Ray	2.1	Assembly of antibody and T cell receptor genes
1E4U	CNOT4	C4C4	NMR	0.58	Part of a global regulator of RNA polymerase II transcription
1G25	Mat1	C3HC4	NMR	0.67	Subunit of the human transcription/DNA repair factor TFIIH
1JM7	BRCA1	C3HC4	NMR	0.87	DNA repair and transcriptional regulation
1JM7	BARD1	C3HC4	NMR	0.95	DNA repair and transcriptional regulation
1FBV	c-Cbl	C3HC4	X-Ray	2.9	Negative regulator of tyrosine kinase-coupled receptors
1LDD/1LDJ	Rbx1	RING-H2	X-Ray	3	Part of the SCF E3 ligase complex

The first X-ray structure of a RING domain was published two years thereafter (Bellon et al., 1997) (Table 1). The structure of the dimerization domain of the protein RAG1 consists of a N-terminal RING finger domain and a C-terminal Zinc finger. The structure of the RING finger domain of RAG1 is highly similar to that of IEEHV (Figure 5). A superposition of 40 of the 44 C α atoms of the RAG1 RING finger onto the corresponding C α atoms of IEEHV finger results in a RMSD of 1.65 Å. The highly conserved hydrophobic core residues Phe309 and Ile314 corresponding to Phe28 and Ile33 residues in the IEEHV adopt identical conformations in the two structures. However, in the structure of RAG1, the third β -strand is missing.

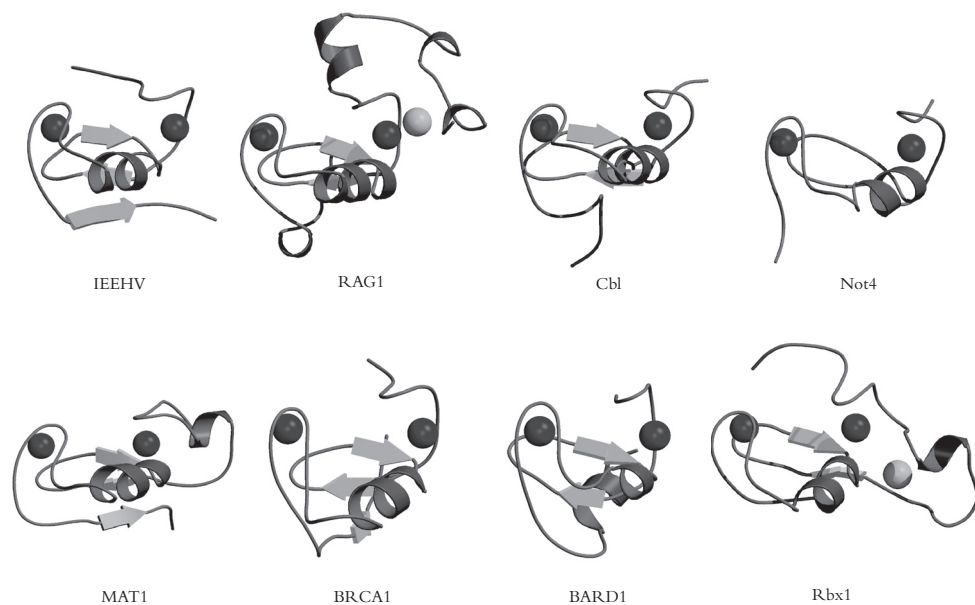


Figure 5: Structures of RING domains: RING finger zinc atoms are represented in black, extra zinc atoms (RAG1 and Rbx1) are represented in grey. The figures have been generated with the program Molscript (Kraulis, 1991) and Raster3D (Merrit and Murphy, 1994).

The structure of a C4C4 type RING domain of the CNOT4 protein from the human CCR4-NOT transcription complex, has been solved using NMR (Hanzawa et al., 2001) (Table 1). Despite the fact that the sequence of the N-terminal domain of CNOT4 did not show the consensus C3HC4 RING motif, the fold of the domain and the zinc coordination topology of CNOT4 resembles that of the RING proteins (Figure 5). The structure of CNOT4 RING domain consists of three loops and a α -helix between the second and third loop. No regular β -strand could be detected in the structure using the Kabsch-Sander algorithm as implemented in the DSSP program. However, despite the differences in secondary structure elements, that are likely due to the presence of many prolines in the CNOT4 sequence, the overall structure of CNOT4 is similar to the structure of IEEHV (C α RMSD: 1.7 Å) and RAG1 (C α RMSD: 1.6 Å). The metal binding site was investigated by replacing the zinc atoms by ^{113}Cd . The ^{113}Cd - ^1H HSQC spectrum showed the presence of two metal binding sites. The cysteines 14, 17, 38 and 41 coordinate one metal ion and the cysteines 31, 33, 53 and 56 bind the second metal ion, in both cases via the SG atoms, confirming the cross-brace motif of the CNOT4 RING domain. The three loops in CNOT4 are stabilized by the coordination with the zinc ions and by hydrophobic interactions in a similar manner as for the C3HC4 RING finger structures. Interestingly, the distance between the two zinc atoms is slightly longer (15 Å) than the conserved value of 14 Å as found in the C3HC4 RING finger structures of IEEHV, and RAG1. This difference

is possible due to the difference in residue spacing between the ligands 4 and 5 of CNOT4 versus those of the C3HC4 RINGs (Figure 2).

Next to CNOT4, three other structures of C3HC4 RING domain were also solved in 2001. One corresponds to the NMR structure of the RING domain of the human transcription factor TFIIF MAT1 (ménage a trois) subunit (Gervais et al., 2001) (Table 1). The structure adopts the $\beta\beta\alpha\beta$ fold that is typical of RING domains. The core of the domain consists of a three-stranded antiparallel β -sheet packed along a two-turn α -helix (Figure 5). The other two structures correspond to the NMR structure of the complex between two RING domains, the BRCA1-BARD1 heterodimeric RING-RING complex (Brzovic et al., 2001) (Table 1). The BRCA1 RING domain is characterized by a short three-stranded antiparallel β -sheet, two large zinc-binding loops and a central α -helix (Figure 5). The BARD1 (BRCA1-associated RING domain protein 1) is structurally homologous but lacks the central helix between the third and fourth pair of zinc ligands (Figure 5). The dimerization interface of BRCA1 and BARD1 does not involve the RING motifs directly as discussed below and the two structures closely resemble those of the free RING domains.

Structures of related domains

A number of domains, such as the PHD domain found in E3 ligases and the FYVE domain found in proteins involved in the membrane recruitment of cytosolic proteins, show the same zinc cross-brace ligation motif as the RING domain (Aasland et al., 1995; Stenmark et al., 2002). The LIM domain found in transcription factors, shows a more conventional sequential ligation pattern (Dawid et al., 1998).

The LIM domain is commonly found in transcription factors, like many Zinc fingers. The domain is involved in protein-protein interaction with other transcription accessory factors. Six structures of LIM domains have been solved thus far, all of them by NMR (Hammarstrom et al., 1996; Konrat et al., 1997; Perez-Alvarado et al., 1996; Perez-Alvarado et al., 1994; Velyvis et al., 2001; Yao et al., 1999). A comparison shows that the LIM domain is less compact than the core structure of the RING domain, probably because of the difference in the zinc ligation topology (Figure 6). The conserved central helix in RING fingers is not present in LIM domains. However, the fold of the LIM domain resembles that of part of the RING domain. A superposition of the structures for the residues around the first zinc (site I) and the two first β -strands leads to a backbone RMSD of 3 Å between the CRP2 LIM domain and the IEEHV RING finger protein.

Initially, the function of proteins containing a PHD domain was unclear, although there was a general consensus that the PHD domains are involved in protein-protein interactions. More recently, PHD-containing proteins have been shown to function in the ubiquitination pathway as E3 ligases. The PHD domain in these proteins has the same function as the RING and the HECT domain, and is involved in the interaction with the E2 ubiquitin conjugating enzyme (Coscoy and Ganem, 2003). The structure of the PHD domains from the human Williams-Beuren Syndrome transcription factor (WSTF) and from the Kap-1 transcriptional repressor have been solved by NMR (Capili et al., 2001; Pascual et al.,

2000). The zinc coordination topology of the PHD domain is similar to that of the RING finger domain, but it turns out that the ternary fold is quite different (Figure 6). The structure of the PHD domain consists of a N-terminal loop followed by an anti-parallel β -sheet, a second loop, and a third β -strand. The site I zinc ligand pairs are located in the N-terminal loop (first pair) and at the beginning of the second loop (third pair). The site II zinc ligand pairs are located in the turn between the two first β -strands (second pair) and at the end of the third β -strand (fourth pair). The arrangement of the site I zinc ligands, as well as the two first β -strands are similar to that of the RING domain (a superposition of these regions between the Kap-1 PHD domain and the IEEHV RING domain leads to a backbone RMSD of 1.3 Å) (Capili et al., 2001). However, the position of the second zinc with respect to the first one is different between the two domains and the conserved α -helix found in the RING domain is not present in the PHD structures.

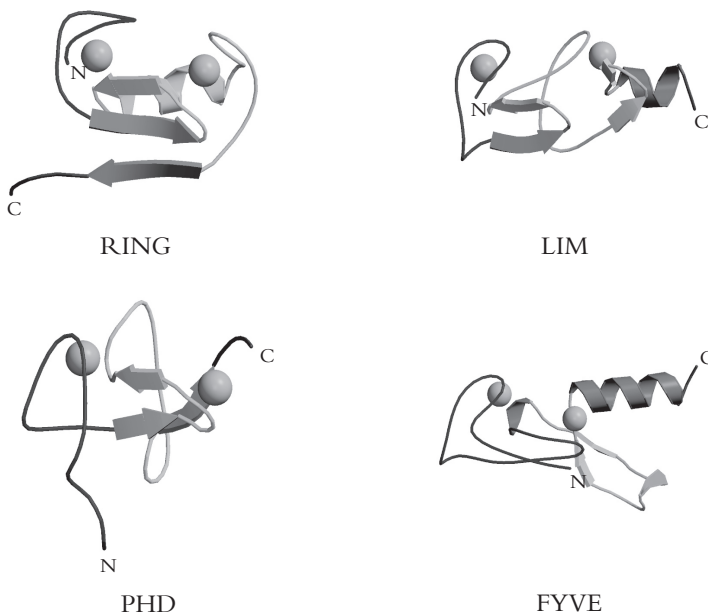


Figure 6: Comparison between the structures of the RING domain of the IEEHV protein, the LIM domain of the CRP2 protein (PDB:1A7I), the PHD domain of the Kap1 protein (PDB:1FPO) and the FYVE domain of the VSP27P protein (PDB:1VFY). The figures have been generated with the program Molscript (Kraulis, 1991) and Raster3D (Merritt and Murphy, 1994).

Also the cross-brace motif found in the FYVE domain is very similar to that of the RING motif. Since the FYVE domain is found in proteins involved in the membrane recruitment of cytosolic proteins and binds to phosphatidylinositol 3-phosphate (PI3P) located in membranes (Stenmark et al., 2002), there is clearly no functional resemblance with the RING and the PHD domains, involved in ubiquitination. Four structures of

FYVE domains have been solved by X-ray (Dumas et al., 2001; Mao et al., 2000; Misra and Hurley, 1999) and NMR (Kutateladze and Overduin, 2001). The FYVE domain is composed of two central β -strands and a C-terminal α -helix (Figure 6). Despite the similarity in zinc coordination topology with the RING and PHD domains, there is no structural relationship between the FYVE and these two domains. The secondary structure elements of the FYVE domain are located at totally different positions as compared to both RING and PHD domains.

Structures of the RING domains in complex with other proteins

Ubiquitination consists of the covalent attachment of ubiquitin molecules to protein substrates. The conjugation of ubiquitin to substrates is accomplished by an enzymatic cascade involving three enzymes (see biological function). How these enzymes recognize each other and what are the molecular basis of these interactions, are key questions to understand the specific mode of action of the ubiquitination pathway. Detailed molecular insights into the E2-E3 interaction came from the crystal structure of E6AP, an E3 protein ligase containing a HECT domain, bound to UbcH7 (human ubiquitin conjugating enzyme 7) an E2 enzyme (Huang et al., 1999). One year later, the crystal structure of the RING finger protein c-Cbl (CAS-BR-M murine ecotropic retroviral transforming sequence homolog) bound to UbcH7 (Zheng et al., 2000) extended our knowledge to the RING-E2 interaction mechanism (Figure 7). E3 protein ligases represent a diverse family of enzymes that can either be single proteins or large protein complexes. The SCF (Skp1, Cullin, F-box protein) complexes, which consist of 4 different proteins, are one of the largest E3 protein ligase complexes. One of the proteins of the SCF complex contains a RING domain that is required for the E2 recognition. How the E3 complex spatially arranges the transfer of ubiquitin from the E2 to the substrate is an important question towards the understanding of the molecular basis of ubiquitination. The crystal structure of the SCF complex composed of the cullin gene family member 1 (Cul1), the RING box protein 1 (Rbx1), the suppressor of kinetochore protein 1 (Skp1), and the F-box domain of the suppressor of kinetochore protein 2 (Fbox^{skp2}) provides considerable insights into this issue (Zheng et al., 2002) (Figure 8).

The c-Cbl-UbcH7 complex

The structure of the full length UbcH7 in complex with the N-terminal half of the 100 kD protein c-Cbl has been solved by X-ray crystallography. The N-terminal half of c-Cbl corresponds to a TKB (tyrosine kinase binding) domain (residues 47 to 344), a linker sequence (residues 345 to 380) and a C3HC4 RING domain (residues 381 to 434) (Zheng et al., 2000). The structure of the RING domain in the complex is similar to that of the free RING domains. The backbone RMSD between c-Cbl and RAG1 for the 40 core residues is 1.9 Å, similar to the RMSD values found between all RING domains (Zheng et al., 2000) (Figure 5). The structure of the complex shows how the RING domain of the E3 ligase interacts with the E2 conjugating enzyme. The interaction involves the two well-conserved zinc-chelating loops and the α -helix of the RING domain (Figure 7). The

interaction is stabilized mainly by hydrophobic contacts, in which the aromatic side chain of Phe63 of UbcH7 plays a crucial role. This residue makes close van der Waals contacts with Ile383 in the loop1 and Trp408, Ser407 and Ser411 in the α -helix of the c-Cbl Ring domain. Other important hydrophobic contacts are between Pro97 and Ala98 of UbcH7 and Ile383, Trp408, Pro417 and Phe418 of c-Cbl. In addition, the charged residues Glu366 and Glu369 of the linker region of c-Cbl and Arg5 and Arg15 of UbcH7 make electrostatic contacts. A comparison of the structure of the c-Cbl-UbcH7 complex and the E6AP-UbcH7 complex reveals considerable similarities in the interaction surface of UbcH7 with these two structurally unrelated E3 ligases. The binding of UbcH7 with both proteins involves the same set of residues, even though, in the RING and the HECT domains, the structural elements and residues, which are recognized by UbcH7, are different. The importance of the loops and α -helix of the RING domain for the interaction with the E2 enzyme are underlined by the NMR studies on the CNOT4-UbcH5B complex (Albert et al., 2002). The NMR studies showed that the regions of the CNOT4 RING domain involved in the interaction with UbcH5B are similar to the ones of c-Cbl in the interaction with UbcH7. However, the types of amino acids involved in the interaction are different, which possibly explains the differences in specificity in the E2-E3 interaction. Whereas in c-Cbl RING domain, the residues Ser407, Trp408 and Ser411 are involved in the interaction with UbcH7, in CNOT4 the corresponding residues Arg44, Ile45 and Asp48 are involved in the interaction with UbcH5B (Figure 2).



Figure 7: The c-Cbl – UbcH7 complex. The RING domain of c-Cbl and the residues important for the interaction (Ile383, Cys384, Cys404, Ser407, Trp408, Ser411, Pro417, Phe418 and Arg420) are represented in black. UbcH7 is represented in light grey. The figure has been generated with the program Molscript (Kraulis, 1991) and Raster3D (Merrit and Murphy, 1994).

The SCF ubiquitin ligase complex

Recently, a second structure of a RING domain containing protein complex has been solved by X-ray (Zheng et al., 2002) (Figure 8). This complex consists of the RING-H2 protein Rbx1 in association with Cul1, Skp1 and Fbox^{skp2} and corresponds to the SCF ubiquitin ligase complex. In this case, the complex provides information about the mechanism of ubiquitin transfer by an E3 ligase. The proteins Rbx1 and Cul1 form the catalytic core of this complex. Rbx1 is responsible for the E2 recruitment. Cul1 consists of a long α -helical structure composed of three cullin repeats and binds Skp1 via its N-terminus and Rbx1 via its C-terminus. The F-box protein Skp2 binds the substrate. Skp1 is an adapter between the F-box protein and Cul1. The protein Rbx1 has a RING-H2 domain that adopts the same fold as that of the canonical RING motif. The RING-H2 domain of Rbx1 is stabilized by two zinc ions, but possess a 20-residue insertion between the first and the second pair of zinc ligand. This insertion contains three additional zinc ligands and together with a fourth zinc ligand from the RING motif, these residues form a new zinc-binding site (Figure 5). The docking of the SCF complex onto the c-Cbl-UbcH7 complex (Zheng et al., 2000) on one side and the Skp1-Skp2 complex (Schulman et al., 2000) on the other (Figure 8) shows that the binding site of Rbx1 for an E2 enzyme is not affected by this additional Zinc finger motif and that the active site Cys86 of UbcH7 and the tip of the F-box protein Skp2 are on the same side of the SCF complex. The distance between the E2 enzyme and the substrate

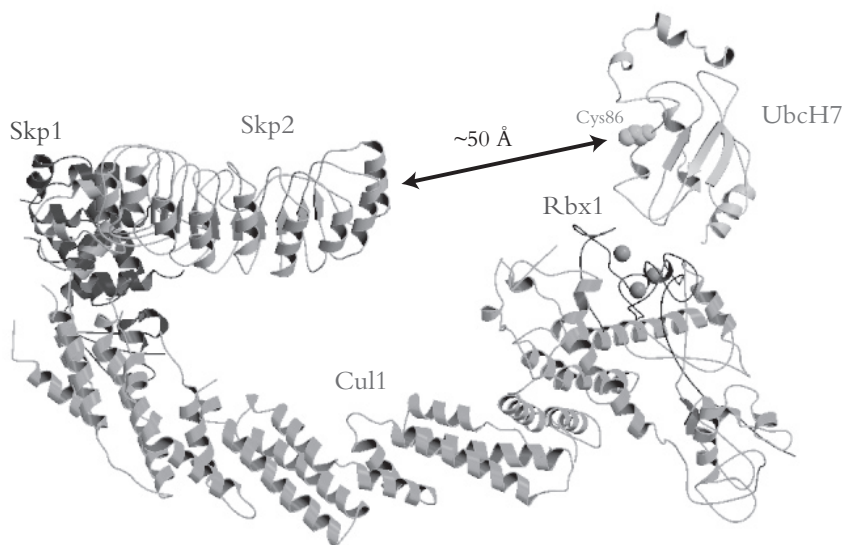


Figure 8: The SCF ubiquitin ligase complex. The X-ray complex consists of the Rbx1 RING, the Cul1, the Skp1 proteins and the Skp2 F-box domain. The UbcH7 structure has been docked on the basis of the c-Cbl-UbcH7 complex (Zheng et al., 2000) as well as the full Skp2 structure on the basis of the Skp1-Skp2 complex (Schulman et al., 2000). The arrow indicates the 50 Å gap between the active site of the E2 (Cys86) and the tip of Skp2. The figure has been generated with the program Molscrip (Kraulis, 1991) and Raster3D (Merrit and Murphy, 1994)

binding site of the F-box protein is about 50 Å. This distance is suitable for the insertion of the substrate, p27 between Skp2 and UbcH7 and for the positioning of p27 at a suitable topology for its ubiquitination.

Another interaction involving RING proteins that has been observed is the direct interaction between two RING domains in the dimeric BRCA1-BARD1 complex (Brzovic et al., 2001). In this heterodimeric complex, the complex formation is mainly governed by extensive interactions involving hydrophobic residues of a four-helix bundle formed between the two proteins, while there are only few direct contacts between the two RING domains.

Medical aspect of RING finger proteins

Many human diseases can be related to a malfunctioning in the regulation of gene expression. As ubiquitination is a key step in the protein degradation pathway, it is not surprising that alterations of this process leads to severe diseases. Since it has been shown that many RING domain proteins function as E3 ligases, there has been a rapid increase in the number of articles that link RING domains with disease. For this review, we will focus on only three RING domain proteins, two involved in cancer and one involved in Parkinson's disease, for which the mode of action is now well defined. The interested reader is referred to recent reviews (Fang et al., 2003; Michael and Oren, 2003; Mizuno et al., 2001; Shtiegman and Yarden, 2003) for more details.

The BRCA1 protein

The gene for the putative tumor suppressor BRCA1 (Breast Cancer 1) was first cloned in 1990 (Hall et al., 1990). BRCA1 encodes a protein of 220 kDa (1863 amino acids) that has a highly conserved amino terminal RING finger and an acidic carboxyl terminal domain characteristic of many transcription factors. BRCA1 has been shown to be involved in several important cellular functions including DNA repair, regulation of transcription, cell-cycle control and ubiquitination (Kerr and Ashworth, 2001). During the last decade, researchers have been able to confidently link mutations in the BRCA1 gene to familial breast and ovarian cancers (Ford et al., 1994; Miki et al., 1994). It was found that mutations in the BRCA1 gene are responsible for about 5 % of the total of these cancers, but it became also clear that a woman that inherits one mutant allele of BRCA1 from either her mother or father has currently a >80 % risk of developing breast cancer during her life. It has been estimated that between 1/500 and 1/800 woman carry a mutation in their BRCA1 gene (Ford et al., 1995). More than 600 different mutations predisposing to a high risk of cancer have been identified in the BRCA1 gene, some of them in the RING domain (Lorick et al., 1999; Serova et al., 1996). In particular mutations of the ligand cysteines 39 (Santarosa et al., 1998), 61 (Friedman et al., 1994) and 64 (Castilla et al., 1994) that probably disrupt the integrity of the RING structure underscore the importance of the RING for functionality of BRCA1.

The p53-Mdm2 complex

Exposure to ionizing radiation or ultraviolet light damages the DNA and leads, in the worst case, to cellular transformation. For maintaining genomic stability the cell is equipped with repair systems, that recognize the DNA damages, initiate repair or direct death of the damaged cells. One of the proteins playing a central part in damage recognition, signal transduction, initiation of apoptosis and repair is the tumor suppressor protein p53 (Levine, 1997; Michael and Oren, 2002). It accumulates to high levels after DNA damage and this increase in abundance, presumably in combination with activating modifications, leads to cell cycle arrest to allow repair of the DNA before the next round of replication or it implements cell death by activating transcription of pro-apoptotic target genes. In normal cells, p53 is degraded by the 26S proteasome. The multiubiquitin-chain is attached to p53 by the onco-protein Mdm2 (murine double minute chromosome clone number 2) (Michael and Oren, 2002). The central domain of Mdm2 is constitutively phosphorylated and this phosphorylation is essential for p53 degradation. Mdm2 possess a carboxyl terminal RING domain (Boddy et al., 1994) that promotes the ubiquitination (Fang et al., 2000), and is necessary for the nuclear exclusion (Boyd et al., 2000; Geyer et al., 2000) and the degradation of p53. It has been estimated that 50% of human cancers are due to mutations of the p53 protein and some of these mutations prevent the Mdm2 mediated degradation (Michael and Oren, 2002).

The Parkin protein

Autosomal recessive juvenile parkinsonism (AR-JP) is one of the most common forms of familial Parkinson's disease and is characterized by selective and massive loss of dopaminergic neurons, leading to a deficiency of dopamine supplies, and absence of Lewy bodies, cytoplasmic inclusions consisting of insoluble protein aggregates. The clinical features of the disease consist of resting tremor, cogwheel rigidity, bradykinesia and postural instability (Tanaka et al., 2001). It is the second most frequent neurodegenerative disorder after Alzheimer's disease. The causative gene of AR-JP is the Parkin gene that encodes a protein of 52 kDa (465 amino acids) that is composed of an amino terminal ubiquitin-like domain and a carboxyl terminal RING-IBR-RING motif. The RING-IBR-RING motif corresponds to two RING domains separated by an additional Cys/His rich domain termed IBR (in between RING) finger (Kitada et al., 1998). The RING-IBR-RING domain is necessary and sufficient for binding UbcH7 (Tanaka et al., 2001). Parkin is involved in protein degradation as an E3 ubiquitin ligase (Shimura et al., 2000) suggesting that dysfunction of the ubiquitin-proteasome pathway plays a role in the Parkinson's disease. It has been shown that Parkin can ubiquitinate itself thereby promoting its own degradation (Zhang et al., 2000). Furthermore, Parkin also induces the degradation of the synaptic vesicle-associated protein, CDCrel-1 (Cell division control related protein 1). Therefore, mutations in Parkin could lead to an accumulation of CDCrel-1 in the brain promoting an inhibition of the release of dopamine (Zhang et al., 2000). Some of the Parkin mutations that are found in the Parkinson's disease involve residues within the RING domains (West et al., 2002). Concerning the RING2, a mutation of the 6th cysteine ligand may disrupt the

integrity of the structure. Interestingly also mutations were found within the RING-1 that could, based on sequence homology with c-Cbl, be involved in the E2 recognition. Further studies should elucidate the mechanism underlying the relationship between the AR-JP form of the familial Parkinson's disease and the deficiency of the ubiquitin ligase activity of Parkin.

Concluding remarks

For many years, the eukaryotic protein degradation pathway by the 26S proteasome, and therefore by the ubiquitination pathway, was considered as a simple protein removal system of the cells. However, it is now clear that the ubiquitination pathway is involved in many regulatory processes. The observation that the RING motif is a key player in the ubiquitination pathway as being an E3 ligase not only shed light on the function of RING finger-containing proteins but also provided an explanation for the diverse biological processes the RING finger proteins participate in. In the last few years, a large body of evidence demonstrated that the ubiquitination pathway mediated by RING domains is an indispensable regulatory process of the cells. This is further underscored by the observation that many diseases, like certain cancers, are due to mutations in RING domains that prevent an efficient ubiquitination and degradation process.

The general dogma, ubiquitination leads to protein degradation by the proteasome, was recently challenged by the observation that (mono)ubiquitination functions as a protein modification that, in the same manor as phosphorylation, acetylation or methylation, regulates cellular location and activity of many proteins involved in transcription, chromatin remodeling or DNA repair. This new twist raises many questions with respect to the function and reaction mechanism of mono-ubiquitination versus poly-ubiquitination and the role of RING fingers herein. Recent structures of protein complexes of E2 conjugating enzymes and E3 ligases provided clear insight into the reaction mechanism of the ubiquitin transfer but many questions remain unanswered. Determination of structures of protein complexes containing ubiquitin conjugating enzymes and ubiquitin ligases together with specific substrates will provide detailed insight into the reaction mechanism of ubiquitin transfer from E2 to the substrates. Such structures could further explain the observed E2-E3 specificities and might clarify the role of ubiquitination in protein activation or destruction and the involvement of RING domains in these processes.

Acknowledgment

The authors are grateful to Sandrine Jayne for critical reading of the manuscript. The financial support of the Center for Biomedical Genetics is also acknowledged.

References

- Aasland, R., Gibson, T. J., and Stewart, A. F. (1995). The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci* 20, 56-59.
- Albert, T. K., Hanzawa, H., Legtenberg, Y. I. A., de Ruwe, M. J., van den Heuvel, F. A. J., Collart, M. A., Boelens, R., and Timmers, H. T. M. (2002). Identification of a ubiquitin-protein ligase subunit within the CCR4-NOT transcription repressor complex. *Embo J* 21, 355-364.
- Aravind, L., and Koonin, E. V. (2000). The U box is a modified RING finger - a common domain in ubiquitination. *Curr Biol* 10, R132-134.
- Barlow, P. N., Luisi, B., Milner, A., Elliott, M., and Everett, R. (1994). Structure of the C3HC4 domain by 1H-nuclear magnetic resonance spectroscopy. A new structural class of zinc-finger. *J Mol Biol* 237, 201-211.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002). The Pfam protein families database. *Nucleic Acids Res* 30, 276-280.
- Bellon, S. F., Rodgers, K. K., Schatz, D. G., Coleman, J. E., and Steitz, T. A. (1997). Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster. *Nat Struct Biol* 4, 586-591.
- Belz, T., Pham, A. D., Beisel, C., Anders, N., Bogin, J., Kwozynski, S., and Sauer, F. (2002). In vitro assays to study protein ubiquitination in transcription. *Methods* 26, 233-244.
- Boddy, M. N., Freemont, P. S., and Borden, K. L. B. (1994). The p53-associated protein MDM2 contains a newly characterized zinc-binding domain called the RING finger. *Trends Biochem Sci* 19, 198-199.
- Borden, K. L. B. (2000). RING domains: master builders of molecular scaffolds? *J Mol Biol* 295, 1103-1112.
- Borden, K. L. B., Boddy, M. N., Lally, J., O'Reilly, N. J., Martin, S., Howe, K., Solomon, E., and Freemont, P. S. (1995). The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *Embo J* 14, 1532-1541.
- Borden, K. L. B., and Freemont, P. S. (1996). The RING finger domain: a recent example of a sequence-structure family. *Curr Opin Struct Biol* 6, 395-401.
- Boyd, S. D., Tsai, K. Y., and Jacks, T. (2000). An intact HDM2 RING-finger domain is required for nuclear exclusion of p53. *Nat Cell Biol* 2, 563-568.
- Brzovic, P. S., Rajagopal, P., Hoyt, D. W., King, M. C., and Klevit, R. E. (2001). Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. *Nat Struct Biol* 8, 833-837.
- Capili, A. D., Schultz, D. C., Rauscher, I. F., and Borden, K. L. B. (2001). Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains. *Embo J* 20, 165-177.
- Castilla, L. H., Couch, F. J., Erdos, M. R., Hoskins, K. F., Calzone, K., Garber, J. E., Boyd, J., Lubin, M. B., Deshano, M. L., Brody, L. C., and et al. (1994). Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nat Genet* 8, 387-391.
- Coscoy, L., and Ganem, D. (2003). PHD domains and E3 ubiquitin ligases: viruses make the connection. *Trends Cell Biol* 13, 7-12.
- Dawid, I. B., Breen, J. J., and Toyama, R. (1998). LIM domains: multiple roles as adapters and

functional modifiers in protein interactions. *Trends Genet* 14, 156-162.

Deng, L., Wang, C., Spencer, E., Yang, L., Braun, A., You, J., Slaughter, C., Pickart, C., and Chen, Z. J. (2000). Activation of the I κ B kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* 103, 351-361.

Dumas, J. J., Merithew, E., Sudharshan, E., Rajamani, D., Hayes, S., Lawe, D., Corvera, S., and Lambright, D. G. (2001). Multivalent endosome targeting by homodimeric EEA1. *Mol Cell* 8, 947-958.

Everett, R. D., Barlow, P., Milner, A., Luisi, B., Orr, A., Hope, G., and Lyon, D. (1993). A novel arrangement of zinc-binding residues and secondary structure in the C3HC4 motif of an alpha herpes virus protein family. *J Mol Biol* 234, 1038-1047.

Fang, S., Jensen, J. P., Ludwig, R. L., Vousden, K. H., and Weissman, A. M. (2000). Mdm2 is a RING finger-dependent ubiquitin protein ligase for itself and p53. *J Biol Chem* 275, 8945-8951.

Fang, S., Lorick, K. L., Jensen, J. P., and Weissman, A. M. (2003). RING finger ubiquitin protein ligases: implications for tumorigenesis, metastasis and for molecular targets in cancer. *Semin Cancer Biol* 13, 5-14.

Folkers, G. E., Hanzawa, H., and Boelens, R. (2001). Zinc Finger Proteins. In *Handbooks on metalloproteins*, I. Bertini, A. Sigel, and H. Sigel, eds. (New York Basel, Marcel Dekker, Inc.), pp. 961-1000.

Ford, D., Easton, D. F., Bishop, D. T., Narod, S. A., and Goldgar, D. E. (1994). Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *Lancet* 343, 692-695.

Ford, D., Easton, D. F., and Peto, J. (1995). Estimates of the gene frequency of BRCA1 and its contribution to breast and ovarian cancer incidence. *Am J Hum Genet* 57, 1457-1462.

Freemont, P. S. (2000). RING for destruction? *Curr Biol* 10, R84-87.

Freemont, P. S., Hanson, I. M., and Trowsdale, J. (1991). A novel cysteine-rich sequence motif. *Cell* 64, 483-484.

Friedman, L. S., Ostermeyer, E. A., Szabo, C. I., Dowd, P., Lynch, E. D., Rowell, S. E., and King, M. C. (1994). Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families. *Nat Genet* 8, 399-404.

Garner, C. D., Hasain, S. S., Bremner, I., and Bordas, J. (1982). An EXAFS study of the zinc sites in sheep liver metallothionein. *J Inorg Biochem* 16, 253-256.

Gervais, V., Busso, D., Wasielewski, E., Poterszman, A., Egly, J. M., Thierry, J. C., and Kieffer, B. (2001). Solution structure of the N-terminal domain of the human TFIIH MAT1 subunit: new insights into the RING finger family. *J Biol Chem* 276, 7457-7464.

Geyer, R. K., Yu, Z. K., and Maki, C. G. (2000). The MDM2 RING-finger domain is required to promote p53 nuclear export. *Nat Cell Biol* 2, 569-573.

Glickman, M. H., and Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev* 82, 373-428.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., and King, M. C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250, 1684-1689.

Hammarstrom, A., Berndt, K. D., Sillard, R., Adermann, K., and Otting, G. (1996). Solution structure of a naturally-occurring zinc-peptide complex demonstrates that the N-terminal zinc-binding module of the Lasp-1 LIM domain is an independent folding unit. *Biochemistry* 35, 12723-12732.

- Hanzawa, H., de Ruwe, M. J., Albert, T. K., van Der Vliet, P. C., Timmers, H. T. M., and Boelens, R. (2001). The structure of the C4C4 ring finger of human NOT4 reveals features distinct from those of C3HC4 RING fingers. *J Biol Chem* 276, 10185-10190.
- Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P. M., Huibregtse, J. M., and Pavletich, N. P. (1999). Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* 286, 1321-1326.
- Jackson, P. K., Eldridge, A. G., Freed, E., Furstenthal, L., Hsu, J. Y., Kaiser, B. K., and Reimann, J. D. R. (2000). The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases. *Trends Cell Biol* 10, 429-439.
- Joazeiro, C. A. P., and Weissman, A. M. (2000). RING finger proteins: mediators of ubiquitin ligase activity. *Cell* 102, 549-552.
- Joazeiro, C. A. P., Wing, S. S., Huang, H., Levenson, J. D., Hunter, T., and Liu, Y. C. (1999). The tyrosine kinase negative regulator c-Cbl as a RING-type, E2-dependent ubiquitin-protein ligase. *Science* 286, 309-312.
- Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577-2637.
- Kaiser, P., Flick, K., Wittenberg, C., and Reed, S. I. (2000). Regulation of transcription by ubiquitination without proteolysis: Cdc34/SCF(Met30)-mediated inactivation of the transcription factor Met4. *Cell* 102, 303-314.
- Kerr, P., and Ashworth, A. (2001). New complexities for BRCA1 and BRCA2. *Curr Biol* 11, R668-676.
- Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y., and Shimizu, N. (1998). Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* 392, 605-608.
- Koegl, M., Hoppe, T., Schlenker, S., Ulrich, H. D., Mayer, T. U., and Jentsch, S. (1999). A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell* 96, 635-644.
- Konrat, R., Weiskirchen, R., Krautler, B., and Bister, K. (1997). Solution structure of the carboxyl-terminal LIM domain from quail cysteine-rich protein CRP2. *J Biol Chem* 272, 12001-12007.
- Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24, 946-950.
- Kuras, L., Rouillon, A., Lee, T., Barbey, R., Tyers, M., and Thomas, D. (2002). Dual regulation of the met4 transcription factor by ubiquitin-dependent degradation and inhibition of promoter recruitment. *Mol Cell* 10, 69-80.
- Kutateladze, T., and Overduin, M. (2001). Structural mechanism of endosome docking by the FYVE domain. *Science* 291, 1793-1796.
- Lai, Z., Freedman, D. A., Levine, A. J., and McLendon, G. L. (1998). Metal and RNA binding properties of the hdm2 RING finger domain. *Biochemistry* 37, 17005-17015.
- Levine, A. J. (1997). p53, the cellular gatekeeper for growth and division. *Cell* 88, 323-331.
- Lorick, K. L., Jensen, J. P., Fang, S., Ong, A. M., Hatakeyama, S., and Weissman, A. M. (1999). RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc Natl Acad Sci U S A* 96, 11364-11369.
- Lovering, R., Hanson, I. M., Borden, K. L. B., Martin, S., O'Reilly, N. J., Evan, G. I., Rahman, D., Pappin, D. J. C., Trowsdale, J., and Freemont, P. S. (1993). Identification and preliminary characterization of a protein motif related to the zinc finger. *Proc Natl Acad Sci U S A* 90, 2112-

2116.

Mallery, D. L., Vandenberg, C. J., and Hiom, K. (2002). Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiquitin chains. *Embo J* 21, 6755-6762.

Mao, Y., Nickitenko, A., Duan, X., Lloyd, T. E., Wu, M. N., Bellen, H., and Quioco, F. A. (2000). Crystal structure of the VHS and FYVE tandem domains of Hrs, a protein involved in membrane trafficking and signal transduction. *Cell* 100, 447-456.

Merrit, E. A., and Murphy, M. E. P. (1994). Raster3D version 2.0: A program for photorealistic molecular graphics. *Acta Cryst D* 50, 869-873.

Michael, D., and Oren, M. (2002). The p53 and Mdm2 families in cancer. *Curr Opin Genet Dev* 12, 53-59.

Michael, D., and Oren, M. (2003). The p53-Mdm2 module and the ubiquitin system. *Semin Cancer Biol* 13, 49-58.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., et al. and Skolnik M. H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266, 66-71.

Misra, S., and Hurley, J. H. (1999). Crystal structure of a phosphatidylinositol 3-phosphate-specific membrane-targeting motif, the FYVE domain of Vps27p. *Cell* 97, 657-666.

Mizuno, Y., Hattori, N., Mori, H., Suzuki, T., and Tanaka, K. (2001). Parkin and Parkinson's disease. *Curr Opin Neurol* 14, 477-482.

Pascual, J., Martinez-Yamout, M., Dyson, H. J., and Wright, P. E. (2000). Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *J Mol Biol* 304, 723-729.

Perez-Alvarado, G. C., Kosa, J. L., Louis, H. A., Beckerle, M. C., Winge, D. R., and Summers, M. F. (1996). Structure of the cysteine-rich intestinal protein, CRIP. *J Mol Biol* 257, 153-174.

Perez-Alvarado, G. C., Miles, C., Michelsen, J. W., Louis, H. A., Winge, D. R., Beckerle, M. C., and Summers, M. F. (1994). Structure of the carboxy-terminal LIM domain from the cysteine rich protein CRP. *Nat Struct Biol* 1, 388-398.

Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu Rev Biochem* 70, 503-533.

Roehm, P. C., and Berg, J. M. (1997). Sequential metal binding by the RING finger domain of BRCA1. *Biochemistry* 36, 10240-10245.

Santarosa, M., Viel, A., Dolcetti, R., Crivellari, D., Magri, M. D., Pizzichetta, M. A., Tibiletti, M. G., Gallo, A., Tumolo, S., Del Tin, L., and Boiocchi, M. (1998). Low incidence of BRCA1 mutations among Italian families with breast and ovarian cancer. *Int J Cancer* 78, 581-586.

Saurin, A. J., Borden, K. L. B., Boddy, M. N., and Freemont, P. S. (1996). Does this have a familiar RING? *Trends Biochem Sci* 21, 208-214.

Scheffner, M., Nuber, U., and Huibregtse, J. M. (1995). Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* 373, 81-83.

Schulman, B. A., Carrano, A. C., Jeffrey, P. D., Bowen, Z., Kinnucan, E. R. E., Finnin, M. S., Elledge, S. J., Harper, J. W., Pagano, M., and Pavletich, N. P. (2000). Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature* 408, 381-386.

Schwabe, J. W. R., and Klug, A. (1994). Zinc mining for protein domains. *Nat Struct Biol* 1, 345-349.

Serova, O., Montagna, M., Torchard, D., Narod, S. A., Tonin, P., Sylla, B., Lynch, H. T., Feunteun, J., and Lenoir, G. M. (1996). A high incidence of BRCA1 mutations in 20 breast-ovarian cancer

- families. *Am J Hum Genet* 58, 42-51.
- Shimura, H., Hattori, N., Kubo, S., Mizuno, Y., Asakawa, S., Minoshima, S., Shimizu, N., Iwai, K., Chiba, T., Tanaka, K., and Suzuki, T. (2000). Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nat Genet* 25, 302-305.
- Shtiegman, K., and Yarden, Y. (2003). The role of ubiquitylation in signaling by growth factors: implications to cancer. *Semin Cancer Biol* 13, 29-40.
- Stenmark, H., Aasland, R., and Driscoll, P. C. (2002). The phosphatidylinositol 3-phosphate-binding FYVE finger. *FEBS Lett* 513, 77-84.
- Swanson, R., Locher, M., and Hochstrasser, M. (2001). A conserved ubiquitin ligase of the nuclear envelope/endoplasmic reticulum that functions in both ER-associated and Matalpha2 repressor degradation. *Genes Dev* 15, 2660-2674.
- Tanaka, K., Suzuki, T., Chiba, T., Shimura, H., Hattori, N., and Mizuno, Y. (2001). Parkin is linked to the ubiquitin pathway. *J Mol Med* 79, 482-494.
- VanDemark, A. P., and Hill, C. P. (2002). Structural basis of ubiquitylation. *Curr Opin Struct Biol* 12, 822-830.
- Velyvis, A., Yang, Y., Wu, C., and Qin, J. (2001). Solution structure of the focal adhesion adaptor PINCH LIM1 domain and characterization of its interaction with the integrin-linked kinase ankyrin repeat domain. *J Biol Chem* 276, 4932-4939.
- von Arnim, A. G., and Deng, X. W. (1993). Ring finger motif of Arabidopsis thaliana COP1 defines a new class of zinc-binding domain. *J Biol Chem* 268, 19626-19631.
- Wang, C., Deng, L., Hong, M., Akkaraju, G. R., Inoue, J., and Chen, Z. J. (2001). TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* 412, 346-351.
- Waterman, H., Levkowitz, G., Alroy, I., and Yarden, Y. (1999). The RING finger of c-Cbl mediates desensitization of the epidermal growth factor receptor. *J Biol Chem* 274, 22151-22154.
- Weissman, A. M. (2001). Themes and variations on ubiquitylation. *Nat Rev Mol Cell Biol* 2, 169-178.
- West, A., Periquet, M., Lincoln, S., Lucking, C. B., Nicholl, D., Bonifati, V., Rawal, N., Gasser, T., Lohmann, E., Deleuze, J. F., et al. (2002). Complex relationship between Parkin mutations and Parkinson disease. *Am J Med Genet* 114, 584-591.
- Yao, X., Perez-Alvarado, G. C., Louis, H. A., Pomies, P., Hatt, C., Summers, M. F., and Beckerle, M. C. (1999). Solution structure of the chicken cysteine-rich protein, CRP1, a double-LIM protein implicated in muscle differentiation. *Biochemistry* 38, 5701-5713.
- Yokouchi, M., Kondo, T., Houghton, A., Bartkiewicz, M., Horne, W. C., Zhang, H., Yoshimura, A., and Baron, R. (1999). Ligand-induced ubiquitination of the epidermal growth factor receptor involves the interaction of the c-Cbl RING finger and UbcH7. *J Biol Chem* 274, 31707-31712.
- Zhang, Y., Gao, J., Chung, K. K. K., Huang, H., Dawson, V. L., and Dawson, T. M. (2000). Parkin functions as an E2-dependent ubiquitin-protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1. *Proc Natl Acad Sci U S A* 97, 13354-13359.
- Zheng, N., Schulman, B. A., Song, L., Miller, J. J., Jeffrey, P. D., Wang, P., Chu, C., Koepp, D. M., Elledge, S. J., Pagano, M., et al. (2002). Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* 416, 703-709.
- Zheng, N., Wang, P., Jeffrey, P. D., and Pavletich, N. P. (2000). Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* 102, 533-539.



**Methodological Introduction:
Modeling of macromolecular interactions
based on NMR data**

For a thorough understanding of the biological function of a macromolecule, knowledge of its three dimensional structure is crucial. The determination of structures of macromolecules is mainly achieved by two methods: X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy. From the statistics of the Protein Data Bank (PDB) (Berman et al., 2000) (<http://www.rcsb.org/pdb/>), almost 20000 X-Ray structures and 3000 NMR structures of free proteins have been solved and deposited to date. Most proteins achieve their function, however, in interaction with other biomolecules such as other proteins, DNA, RNA, lipids or sugars. Although a number of methods are available to study the formation of macromolecular complexes (two-hybrid screening, fluorescence spectroscopy, mass spectrometry, surface plasmon resonance, calorimetry, etc), only few techniques provide information at an atomic level. X-ray crystallography provides a detailed atomic picture of molecules, when arranged in a crystalline lattice, whereas NMR spectroscopy provides structural information of biomolecules in solution. Both methods, however, have their shortcomings, especially when applied to studies on biomolecular complexes. The difficulty with crystallography is the co-crystallization of the two macromolecules, especially when the complexes have weak association constants and exist only transiently, whereas for NMR spectroscopy the size of the resulting complex can be a severe limitation. Because of these limitations, the number of macromolecular complexes solved and deposited in the PDB is rather low (less than 1000 by X-Ray and 100 by NMR) as compared to the number of structures in the free form.

NMR structure of macromolecular complexes

The traditional approach to solve macromolecular complexes by NMR spectroscopy requires the collection of proton–proton NOEs (Nuclear Overhauser Effects), which can be translated into distances. When the structures of the separate biomolecules are already known, a structure of the complex can be build up by selectively analyzing the intermolecular NOEs. In addition to these intermolecular NOEs, also orientational restraints like residual dipolar couplings (RDCs), heteronuclear relaxation rates, or paramagnetic pseudo-contact shifts can be used, that precisely define the orientation of the two macromolecules in the complex. These methods require, however, extensive series of experiments and detailed analysis. Another way of studying complex formation by NMR is chemical shift perturbation (CSP) mapping. In this procedure, the chemical shifts changes of given atoms are monitored in the course of the formation of the complex. This procedure is very straightforward and is widely used to map macromolecule–macromolecule interfaces. The information obtained by CSP analysis, however, is only qualitative and does not provide precise distance or orientation information, and has therefore not been commonly used in structure calculations.

Intermolecular Nuclear Overhauser Effect

The main structural information for structure calculations of macromolecules by NMR are distances obtained from NOEs (Wüthrich, 1986). The NOE originates from dipolar cross-relaxation between protons and depends on the proton–proton distance and the molecular motions of the interproton vector. An approximate relation for the NOE intensity is:

$$\text{NOE} \propto 1 / r^6 > f(\tau_c)$$

where r is the proton–proton distance and f a function that depends, among others, on the correlation time (τ_c) that describes the motion of the interproton vector. For rigid molecules, the values of all proton–proton correlation times are defined by the overall tumbling of the molecule, which in the simplest case can be described by a single isotropic overall rotational correlation time. In that case, ratios of NOE intensities provide ratios of distances. These NOE intensities can be easily measured in 2D and 3D NOE spectra. Internal mobility attenuates the NOE effect in biomolecules and therefore the NOEs are generally translated into an upper bound to the distance rather than into a precise distance. In practice the NOE becomes very weak at distances larger than 5–6 Å and often these weak NOEs contain contributions from spin diffusion and should be used with care. Therefore, only an incomplete set of relatively imprecise distance restraints can be used in structure calculations, which then leads to a series of possible solutions. The more precise distance restraints and the larger number of distance restraints are obtained, the closer these solutions will be. NMR structures are therefore generally presented as an overlay of wireframes, where a close overlay represents the well determined regions.

To solve the structure of a macromolecular complex by NMR, structural information is needed in addition to the intramolecular NOEs to define the contacts between the two molecules. The most reliable information that can be obtained comes from proton–proton distances derived from intermolecular NOEs. If the lifetime of the complex is long enough (i.e. a complex with a low dissociation rate), it is possible to observe clear NOEs between protons that belong to different molecules. The simplest way to obtain these intermolecular NOEs is to record a homonuclear NOESY spectrum of a complex and to search for the additional NOEs as compared to NOESY spectra of the free biomolecules. Analysis of such spectra, however, is a lengthy process and it is often difficult to discriminate the intra and intermolecular NOEs. The use of isotope labeling and filtering can overcome this problem. The introduction of ^{13}C and ^{15}N isotopes into one of the biomolecules of the complex allows the use of isotope–filtered experiments that unambiguously assign intermolecular proton–proton NOEs (reviewed in Breeze, 2000). Isotope–filtered experiments have been first introduced by Otting and Wüthrich, (1989) as two–dimensional X–double–half–filtered experiments, where X stand for the “NMR–active” heteroatom (^{15}N and ^{13}C for proteins). In addition, many hybrid schemes combining purge filters and 3D X–separated NOESY spectra have been developed to allow the recording of three–dimensional spectra (Lee et al., 1994; Ogura et al., 1996; Zwahlen et al., 1997). These isotope–filtering techniques rely on the presence of a scalar coupling between a proton and the attached “NMR–active” heteroatom (^{13}C for example) and the absence of a scalar coupling between a proton and an

“NMR-inactive” heteroatom (^{12}C for example). The scalar coupling can be exploited to discriminate between these two different classes of protons. The classical scheme to measure intermolecular NOEs is to study complexes in which one protein is isotopically labeled and the other one is unlabeled. The experiment corresponds to a ^1H - ^1H NOE experiment that is isotope-filtered in one proton dimension and isotope-edited in the other proton dimension. This results in a spectrum containing exclusively intermolecular NOEs providing that the scalar coupling is similar for all proton-heteronuclei pairs and that the labeled protein contains a high percentage of the “NMR-active” isotope (at least more than 98%).

The study of macromolecular complexes by NMR has been facilitated enormously by these isotope-filtered NOE experiments. A limitation is that the additional time in the isotope-filtered experiments leads to lower sensitivity in the NOESY spectra and therefore only an incomplete set of NOEs can be obtained by these methods. In most cases a conventional analysis of unfiltered NOESY spectra is also necessary.

Orientalional restraints

Residual dipolar couplings

The use of residual dipolar couplings (RDCs) in protein solution NMR has been described for the first time in 1995 (Tolman et al., 1995). The magnetic dipole-dipole coupling between the atomic magnetic moments of two nuclei (^{15}N - ^1H for example) depends on the distance between the nuclei, the orientation of the vector defined by the two nuclei with respect to the magnetic field, and the dynamics of this vector. In solution, these dipolar couplings are generally averaged to zero because of the averaged randomly orientated distribution of the interactions. When a slight orientational preference is induced in this distribution due to partial alignment of the macromolecules, the dipolar couplings will not average anymore to zero resulting in residual dipolar couplings. The induced magnetism in macromolecules is generally too weak to lead to observable RDCs. Only in the case of paramagnetic molecules and systems with large inducible magnetic moments such as DNA, is the alignment sufficiently strong to lead to measurable effects. It was discovered, however, that molecules that are dissolved in magnetically aligned crystalline media also obtain a slight preferential orientation. Bicelles-containing solutions were the first liquid crystalline media that have been reported to result in large RDC effects for proteins in solution (Tjandra and Bax, 1997). The degree of this alignment that is due to orienting collisions with the aligned bicelles, depends on the bicelle concentration, and this gives the possibility to control the strength of the dipolar couplings to be high enough to be measurable and not too high to cause excessive line broadening due to unresolved ^1H - ^1H dipolar couplings. After the discovery of the bicelles, many media have been described to align macromolecules such as bacteriophages, silicate rods, polyacrylamide gels and Helfrich phase media (reviewed in Bax, 2003; Bax et al., 2001). These residual dipolar couplings are generally measured for one- and two-bonds interactions since in that case the internuclear distance is defined and the RDC depends mainly on orientation. The most common and straightforward coupling

that can be measured is the $^1D_{\text{NH}}$ dipolar coupling but weaker couplings such as $^1D_{\text{C}\alpha\text{C}}$, $^1D_{\text{C}\alpha\text{N}}$ or $^2D_{\text{C}\alpha\text{HN}}$ can also be measured. Dipolar couplings contain information on the orientation of inter-nuclear vectors relative to a molecular alignment axis and therefore restraint the orientations of these vectors relative to a common frame, the alignment tensor.

Diffusion anisotropy

Analysis of R_1 , R_2 and heteronuclear NOE rates of amide ^{15}N nuclei in the backbone of a protein allows to characterize the motional behavior of the protein backbone in high detail (Kay et al., 1989; Peng and Wagner, 1994). Whereas the R_1 , R_2 and heteronuclear NOE rates reflect the internal dynamics, it has been noted that the ratio R_2/R_1 , at least when exchange broadening can be excluded, are much less sensitive for internal motions and mainly reflect the rotational correlation time of the N-H vector. In the case of a rigid anisotropic molecule, this R_2/R_1 ratio therefore defines the orientation of this vector with respect to an anisotropic diffusion tensor. Therefore, the R_2/R_1 ratios can be translated into diffusion anisotropy restraints to define the orientation of the N-H bonds (Tjandra et al., 1997).

Orientalional restraints for structure calculations

Information derived from residual dipolar couplings and diffusion anisotropy can be used to define orientational restraints that can be used in structure calculations. In the case of macromolecular complexes, these restraints allow to orient one protein with respect to the other one, reducing considerably the number of possible solutions for the structure of a complex. The use of RDCs or diffusion anisotropy for orienting the components of a complex is based on the assumption that the macromolecular complex is more or less rigid and that the two molecules in the complex have a common alignment or diffusion tensor. In that case, one can determine a tensor of each individual component of the complex and rotate them in a way that the two tensors coincide. Though RDCs can be used alone to define macromolecular structures (Hus et al., 2001), orientational restraints are frequently combined with other NMR restraints, such as NOEs. The combined use of these restraints has large advantages: the conventional restraints (NOE and scalar couplings) define well the local geometry while the RDC and the diffusion anisotropy are particularly sensitive to a global relative position of structured regions, especially in the case of macromolecular complexes (Garrett et al., 1999).

Chemical shift perturbation mapping

Chemical shift perturbation mapping obtained via NMR titration experiment is a straightforward NMR technique to define the residues of a protein involved in binding to a partner molecule (reviewed in Zuiderweg, 2002). This type of experiment can already be performed early in the NMR analysis, at the stage of backbone assignments, and allows to identify possible amino acid residues that are involved in complex formation. Its principle relies on the fact that the chemical shift of a nucleus is sensitive to its environment. The

chemical shift changes are measured in a series of NMR experiments (generally ^{15}N HSQC spectra) of the free protein and of the protein with increasing amounts of the partner molecule such as ligand, protein, and DNA. The proximity of the partner in the complex will modify the environment of the nuclei that are at the interface of the complex. As a result, nuclei involved in the binding will have a different chemical shift than in the unbound form. Chemical shift perturbations are generally measured on backbone amides for a number of reasons. First, only ^{15}N -labeled samples are needed and they can be easily obtained by protein expression in *E. coli*. Next, measuring on backbone atoms implies that no side chains assignments are required. Finally, the ^{15}N and ^1H chemical shifts of the amide protons turn out to be sensitive to environmental changes, partially due to subtle changes in the polarization of hydrogen bonds. It is clear that the observed chemical shift changes upon binding can also result from indirect conformational changes in the protein due to the binding. Cases have been reported in which a protein will completely change its conformation upon complex formation; in that case, all chemical shifts are therefore perturbed. In most cases, however, complex formation only induces few or minor structural rearrangements and chemical shift perturbation can then be used to define the interface residues. This technique, however, is generally applied only qualitatively and does not allow to unambiguously identify pairs of atoms that are in contact in the complex or to get precise intermolecular distance information. Furthermore, it does not provide any information about the orientation of one protein with respect to its counterpart. Because of that, chemical shift perturbation information results have rarely been used directly for structure calculations.

Ab Initio docking

Next to experimental approaches, theoretical methods are now emerging to study macromolecular complexes at a structural level based on docking. Some of these methods have been well developed only during the past few years and the methodology is not yet as widely established as X-ray crystallography and NMR spectroscopy. Docking embraces computational methods, generally based on theoretical knowledge, with the task of predicting the assembly of two separate (bio)molecules into their biologically relevant complex. There are a number of programs performing *ab-initio* macromolecular docking (without knowledge on the binding properties) mainly focusing on protein-ligand and protein-protein complexes (reviewed in Camacho and Vajda, 2002; Halperin et al., 2002; Smith and Sternberg, 2002). Most of these programs use the same approach: one protein is fixed in space and the second one is rotated and translated around the first one. For each new configuration, a score is calculated based on various terms such as surface complementarities, electrostatic interactions, van der Waals repulsion.

Docking protocols can be divided into three main sections:

- The representation of the system and the introduction of flexibility

- The search procedure
- The scoring function

The representation of the system depends on the definition of the degrees of freedom for the search procedure. There are different approaches to represent the system (see below). An important point in the representation of the system is the introduction of some flexibility at the interface to account for possible structural rearrangements that can take place during binding. The search procedure consists of the algorithm used to dock the two molecules together (translation, rotations, ...). Two important elements in a search procedure are speed and effectiveness in covering the relevant conformational space. The scoring function is based on a function (often expressed in terms of energies) that should discriminate the various generated solutions. The evaluation of this function should be fast enough to allow its application to a large number of potential solutions and, in principle, the function should effectively discriminate between native and non-native docked conformations. These three main aspects of docking will be discussed in more details in the following.

Representation of the system and introduction of flexibility

There are two main ways of representing the system, geometrically or chemically. A chemical representation can include all atoms, a selected subset or a simplified representation of amino acids. Most docking approaches use, however, a surface representation of the molecules. The surface is usually represented by its geometric features such as, for example, a Connolly surface that consists of the part of the van der Waals surface of the exposed atoms and that is connected by a network of convex or concave shape surfaces (Connolly, 1983). To align the surfaces of two molecules, one needs to superimpose them without allowing one molecule to deeply penetrate or overlap with the other. This is usually performed by aligning triplets of ordered non-collinear points that describe the molecular surface from both molecules. These points should be computed to accurately represent the maxima (holes) and minima (knobs) of the shape function (Norel et al., 1994; Norel et al., 1999). For docking, one of the molecules is translated and rotated, while the other is kept fixed, such as the surface complementarity (represented by the triplets of points) is maximum, i.e. so that the knobs of one molecule face the holes of the other. Physico-chemical features of the molecular surface such as electrostatics can also be added into the geometrical description (Nicholls et al., 1993; Ritchie and Kemp, 2000).

Introduction of flexibility

When docking unbound structures of the molecules, one should account for a certain degree of flexibility at the interface to allow for possible structural rearrangements. The complexity and computational cost of docking procedures depends largely on the extent of flexibility introduced. Docking approaches can be classified into three levels:

- Rigid body docking
- Semi-flexible docking
- Flexible docking

Rigid body docking is the simplest because the proteins are considered as two rigid bodies. Semi-flexible docking is typically asymmetric and mainly used in protein–ligand docking: the ligand is then considered flexible, while the protein is kept rigid. Flexible docking treats both molecules as flexible but the extent of flexibility that is allowed is limited or simplified. Flexibility in docking algorithms has mainly been introduced for protein–ligand complexes for two main reasons. First, ligands are small molecules that are likely to undergo conformational changes upon complex formation. Second, the small size of most ligands allows computationally affordable flexibility algorithms. Attempts to introduce flexibility into proteins have also been described. The two main ways of introducing flexibility in protein–protein docking, which are not computationally expensive, are to allow intermolecular penetration in the search procedure (Gardiner et al., 2001; Jiang and Kim, 1991) or to use an ensemble of starting structures (Knegtel et al., 1997). Such ensembles can be taken from an NMR ensemble, or an ensemble of X-ray structures of the same molecule if available. Otherwise, starting structures can be generated using random thermodynamic sampling (Totrov and Abagyan, 1997), genetic algorithms (Blommers et al., 1992; Legrand and Merz, 1992; Payne and Glen, 1993), or molecular dynamic or monte-carlo simulations (Clarage et al., 1995).

The search procedure

One of the *ab initio* docking problems resides in the many possible ways to put the two molecules together. There are, indeed, three translational and three rotational degrees of freedom in the case of a rigid body docking. Introduction of flexibility increases the number of degrees of freedom, the search space, and thus the computational costs. There are two main approaches to the search problem in docking: a full search of the solution space or a gradual, guided progression through it. The first approach scans the entire solution space by systematically rotating and translating one molecule with respect to the other, while the second scans only part of the solution space in a partially random, and partially criteria-guided manner (see the scoring function section). The second approach is generally based on Monte-Carlo, molecular dynamics, or genetic algorithms. It has been shown that, in general, molecular dynamics algorithms are the most efficient and genetic algorithms the least (Vieth et al., 1998). In the case of protein–protein docking, introduction of flexibility increases dramatically the number of degrees of freedom and full conformational space search becomes impractical. This is why *ab initio* docking approaches applied to protein complexes treat most often the molecules as rigid entities.

The scoring function

Search algorithms typically produce a very large number of possible solutions. The challenging step in docking is to identify the native or near-native solutions using scoring functions within a reasonable computation time. The scoring of solutions is the major issue in docking. In most approaches, a correct solution is ranked within the one hundred to one

thousand best solutions. In most cases, however, the highest ranked solutions are often false positives. This problem is mainly due to the fact that native complexes do not necessarily possess the largest buried surface area, the largest number of hydrogen bonds or the largest non-polar buried surface (Norel et al., 1999). The current solution to overcome this problem is a two-stage approach in which, first one uses a rapid scoring to predict “good” candidates, and second, one uses a more advanced scoring function to discriminate between the selected candidates (Hoffmann et al., 1999).

The scoring approaches can be divided into two groups depending on the stage at which scoring is introduced in the search algorithm. *Integrated approaches* integrate scoring into the search procedure and filter emerging solutions, while *edge approaches* apply scoring functions at the end of the search procedure. Some algorithms, such as genetic algorithms, require integrated approaches.

Most docking programs use similar scoring parameters. The most common are geometric complementarity, intermolecular overlap, number of hydrogen bonds, and number of unsatisfied buried charges, although other parameters based on electrostatic potentials, van der Waals energies, or solvation energies have also been described. Docking programs, however, differ in the way these parameters are implemented in the search algorithms. Geometric complementarity criteria were the first parameters used in docking. Many early docking programs used exclusively geometric complementarity as scoring function. These scoring functions have been quite successful in “bound” docking, that is when the starting structures correspond to the bound form of the molecules. In “unbound” docking, however, geometric complementarity is not sufficient to discriminate between correct and incorrect solutions and the use of other criteria in combination with geometric complementarity is thus necessary (Norel et al., 1999). In these cases, geometric complementarity is used as a primary filter before introducing other more computationally costly evaluation criteria. The degree of intermolecular overlap is then computed and typically penalized in the scoring. A way of dealing with flexibility, however, is to allow intermolecular overlap in a first stage of the docking and then remove it in a refinement step (Gardiner et al., 2001; Palma et al., 2000). In addition to shape complementarity (geometric complementarity and intermolecular overlap), hydrogen bonding or electrostatic criteria can be used. Hydrogen atoms can be added to crystallographic structures according to standard geometry criteria (Barlow and Thornton, 1983; Vriend, 1990). Hydrogen bond formation potentials usually consider four classes of atoms, the H-bond donors, the H-bond acceptors, the H-bond donors/acceptors, and the non-H bonding atoms (Ausiello et al., 1997; Gardiner et al., 2001; Jiang and Kim, 1991). The various approaches can differ in their definition of donors and acceptors, in the introduction of discrimination between bond types, or in the H-bond distance and angular cut-offs used. Electrostatic potentials based on the Poisson-Boltzman equation have also been used in scoring schemes (Gabb et al., 1997; Palma et al., 2000). Another scoring criterion is the buried surface area or contact area. Based on an analysis of known protein-protein complexes, it has been estimated that the average buried surface area of a complex is $1600 \pm 400 \text{ \AA}^2$ (Lo Conte et al., 1999). This criterion correlates with the hydrophobicity at the interface. Many scoring schemes use a hydrophobicity factor (Norel et

al., 1999) or a buried surface area criterion (Gardiner et al., 2001). Solvation energies have also been described in some cases (Camacho et al., 2000; Jackson et al., 1998).

Scoring of solutions remains the most difficult issue in docking. A possible solution is to combine several scoring schemes and derive a consensus scoring in order to reduce the number of false positives (Bissantz et al., 2000; Charifson et al., 1999; Terp et al., 2001).

Protein–protein docking based on NMR data

The drawback of *ab initio* docking is that the search through the entire conformational space for the geometry of the complex makes the calculation time consuming and rarely resulting in a unique solution. The inclusion of experimental data into docking was early described by Weber et al., (1992) and a similar approach was followed by Knegtel et al., (1994b) in docking studies on protein–DNA complexes. In this study, three protein–DNA complexes are docked using MONTY (Knegtel et al., 1994a) based on biochemical information such as mutagenesis experiments. Recently, NMR data have been used in different ways in combination with docking methods to calculate protein–protein complexes.

An early example of a protein–protein docking based on NMR data made use of chemical shift perturbation mapping and intermolecular pseudocontact shifts combined with restrained rigid-body molecular dynamics to solve the structure of the paramagnetic plastocyanin–cytochrome f complex (Ubbink et al., 1998). Measurement of pseudo-contact shifts is, however, limited to paramagnetic proteins.

Later, intermolecular NOEs and residual dipolar couplings (RDCs) have been combined with rigid-body energy minimization to solve the structure of the EIN–HPr complex starting from the structures of the two free proteins (Clare, 2000). This method, however, relies on the collection of intermolecular NOEs and residual dipolar couplings that can be a lengthy and tedious process (see above).

The same year, Morelli *et al.* used the *ab initio* docking program BIGGER (Palma et al., 2000) incorporating a NMR filter, based on chemical shift perturbation data, to select the experimentally relevant solutions (Morelli et al., 2000; Morelli et al., 2001). The approach was tested on four complexes and the best solutions had backbone RMSDs between 0.5 and 2.2 Å from the original structures. This approach allows the use of NMR titration data to rank the possible solutions but the docking is not directly driven by these data and thus still depends on the performance of the search algorithm, i.e. the correct solution should be present in the ensemble of solutions before applying the NMR filter.

Another protein–protein docking approach based on unassigned one-dimensional ¹H NMR spectra of complexes was described (Kohlbacher et al., 2001). In this case, a rigid-body protocol is used to generate an ensemble of complexes in a similar manner as in classical *ab initio* docking. The originality of the method resides in a new scoring scheme. For each generated solution, a theoretical one-dimensional spectrum is calculated based on statistical and theoretical knowledge such as random coil shift of specific atoms, magnetic anisotropy of the peptide bond, ring current effect and electric field effect. This calculated

spectrum is then compared to the experimental one and a difference area (difference between the experimental and the calculated spectra) term is calculated and added to the scoring function. This approach was tested on four complexes. For three of them, the best solution of the docking was the closest to the original structure of the complex (backbone RMSD between 2 and 5 Å).

Recently, a new docking program, TreeDock, was developed where the docking is driven by anchors points, which can be in principle derived from NMR chemical shift perturbation or mutagenesis data (Fahmy and Wagner, 2002). This program performs a rigid-body docking and the solutions are ranked in function of their Lennard-Jones potentials. The approach has been tested on three known complexes starting from the bound structure of each molecule. The backbone RMSD between the best solutions and the structures of the complexes varied between 0.3 and 0.7 Å.

McCoy *et al.* used chemical shift perturbation data in combination with RDCs to develop a new docking approach (McCoy and Wyss, 2002), which was tested on the EIN-HP α complex starting from the unbound forms of the molecules. In that case, the RDC data were first introduced to orient the complex. The solutions were then optimized by back calculating the chemical shift perturbation with SHIFTS (Xu and Case, 2001) and comparing them with the experimental data. The final solutions have a backbone RMSD of approximately 2.5 Å from the original structure.

Later, another docking approach was published using a filtering scheme based on experimental data (Dobrodumov and Gronenborn, 2003). In a first step, *ab initio* rigid-body docking is performed using the 3D-DOCK package (Moont *et al.*, March 2001). The scoring function was modified to incorporate sorting routines based on dipolar couplings (Q factor term), on explicit C α -C α distances that can be derived from biological information, or on mapped contacts derived from NMR chemical shift perturbation data. The method was tested on two complexes for which the structures of the free components and of the complex were available. The best ranked structures display a backbone RMSD between 2.0 and 2.5 Å from the original structures. As is the case in BIGGER (Morelli *et al.*, 2000), this approach, however, only uses NMR data *a posteriori* to rank the possible solutions and depends therefore on the accuracy of the *ab initio* search algorithm.

All these methods based on NMR data consist of rigid-body docking algorithm. During complex formation, however, some structural rearrangements usually occur. As described above, it is important, when docking two proteins, to allow for side chains and/or backbone flexibility in order to find the best orientation of the side chains that will lead to the minimum energy and the best intermolecular contacts. For this, the side chains at the interface should be free to move to adopt their preferential conformation.

We developed a docking protocol that uses directly NMR data to drive the docking process and allows for side chain and backbone flexibility at the interface (Dominguez *et al.*, 2003). This approach is described in Chapter 3 of this thesis. NMR or mutagenesis data are directly used as ambiguous distance restraints. HADDOCK (High Ambiguity Driven DOCKing) makes use of CNS (Brunger *et al.*, 1998) for structure calculation and of Python scripts derived from ARIA for automation (Linge *et al.*, 2001). The docking

protocol allows for side chains and backbone flexibility at the interface. The solutions are scored according to an intermolecular energy term, i.e. sum of electrostatic, van der Waals and AIR (Ambiguous Interaction restraint, see Chapter 3) energy terms, although buried surface area can also be included. This approach has been tested on three complexes starting from both bound and unbound structures. The RMSD between the best solutions and the original structures varied between 0.8 and 2.0 Å. This approach has been further improved by using ensembles of starting structures instead of a single structure as described in Chapter 5 of this thesis (Dominguez et al., 2004).

A very similar approach using highly ambiguous restraints was described (Clore and Schwieters, 2003) shortly after HADDOCK has been published. In addition to the ambiguous interaction restraints derived from chemical shift perturbation mapping as we used, Clore and Schwieters also used RDC data in the calculation. The main differences between the two methods reside in the upper bound limit of the AIR restraints (5 Å instead of 2-3 Å in HADDOCK) and in the final scoring of the solutions. While, in HADDOCK, the solutions are ranked according to an intermolecular energy term (see above and Chapter 3), in the other case, the solutions are ranked based on the ambiguous distance restraint violations and on the R factor of the RDCs. This approach has been tested on three complexes and the backbone RMSD between the best solutions and the original structures is between 0.5 and 1.5 Å.

CAPRI (Critical assessment of PRedicted Interactions)

In order to assess the capacity of current docking methods to predict protein-protein interactions, CAPRI, a community-wide experiment, was set up as a blind test of docking (Janin et al., 2003), similarly to the homology modeling blind test CASP (Critical Assessment of methods for protein Structure Prediction) (Venclovas et al., 2003). CAPRI relies on the generosity of experimentalists willing to communicate unpublished atomic coordinates of protein-protein complexes on a confidential basis. CAPRI started in 2001 with three targets and 19 predictor groups. Since then, 5 rounds of CAPRI have been organized (round 5 was in progress at the time of writing of this thesis). The various evaluation criteria include the RMSD between the predicted and the experimental complex and the number of native intermolecular contacts present in the predicted models (Mendez et al., 2003). Evaluation of the CAPRI predictions and of the performance of the various docking approaches provides useful insights for future developments and improvements of these methods. The CAPRI results until now indicate that a major problem is the scoring of solutions. It has become clear that the use of experimental information is an important element to discriminate between the various solutions. Using HADDOCK, we have participated in round 4 of CAPRI. Our approach, HADDOCK, allowed the use of a wide range of different experimental data to drive the docking (mutagenesis, conservation of exposed residues, epitope mapping, etc). HADDOCK appeared to be very successful in this CAPRI round where our predictions ranked at the top for three out of four targets. In particular, we successfully predicted the trimeric form of a

flavivirus envelope glycoprotein to within 2.8 Å of the experimental model that was recently published (Bressanelli et al., 2004). This was basically the only correct solution out of the 190 predictions that were deposited.

Concluding remarks

In the current structural genomics era, massive efforts are made to solve macromolecular structures. Both X-ray crystallography and NMR are developing new methodologies to automate and speed up the structure determination process. These efforts are currently mainly focusing on structures of single proteins or domains thereof. Most proteins, however, function by interacting with partner molecules that can be another protein, DNA, RNA, sugar or lipids. The understanding of protein function highly depends on our knowledge of the interactions that are taking place and thus, there is a clear need for models of biomolecular complexes. Information on such molecular interactions is also crucial for pharmaceutical applications such as drug design.

Ab initio docking appears to be a complementary approach to X-ray crystallography and NMR spectroscopy for predicting structures of complexes in cases where these two methods fail. Many improvements have been proposed over the last few years and these methods are getting more and more robust. *Ab initio* docking, however, still suffers from important limitations. First, during the search procedure, one should cover the entire conformational space. Second and most important, scoring functions need to be improved to discriminate between the various solutions from the huge amount of solutions that are generated. Third, introduction of flexibility at the interface is crucial but is computationally very time consuming, and therefore fully flexible docking is rarely used.

During the last five years, structural biology and theoretical docking approaches have been combined to predict more accurately structures of macromolecular complexes. The use of NMR data, and also biochemical data, has been introduced into docking algorithms and scoring functions. The experimental data optimize the scoring of the solutions and secondly reduce the search in the vast conformational space. If the binding interface of a macromolecule is known, it is not necessary to generate solutions that do not present this interface. The advantage of reducing the space search is that it considerably reduces the computation time, allowing for docking algorithms to be optimized to include flexibility at the interface. These new methods are clearly limited by the quantity and quality of biochemical and/or biophysical information that is needed to generate realistic solutions. This new field of information-driven docking is, however, just emerging and will probably become even more powerful in a near future.

References

- Ausiello, G., Cesareni, G., and Helmer-Citterich, M. (1997). ESCHER: a new docking procedure applied to the reconstruction of protein tertiary structure. *Proteins* 28, 556-567.
- Barlow, D. J., and Thornton, J. M. (1983). Ion-pairs in proteins. *J Mol Biol* 168, 867-885.
- Bax, A. (2003). Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci* 12, 1-16.
- Bax, A., Kontaxis, G., and Tjandra, N. (2001). Dipolar couplings in macromolecular structure determination. *Methods Enzymol* 339, 127-174.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res* 28, 235-242.
- Bissantz, C., Folkers, G., and Rognan, D. (2000). Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* 43, 4759-4767.
- Blommers, M. J., Lucasius, C. B., Kateman, G., and Kaptein, R. (1992). Conformational analysis of a dinucleotide photodimer with the aid of the genetic algorithm. *Biopolymers* 32, 45-52.
- Breeze, A. L. (2000). Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog Nucl Mag Res Sp* 36, 323-372.
- Bressanelli, S., Stiasny, K., Allison, S. L., Stura, E. A., Duquerroy, S., Lescar, J., Heinz, F. X., and Rey, F. A. (2004). Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *Embo J* 23, 728-738.
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54 (Pt 5), 905-921.
- Camacho, C. J., Gatchell, D. W., Kimura, S. R., and Vajda, S. (2000). Scoring docked conformations generated by rigid-body protein-protein docking. *Proteins* 40, 525-537.
- Camacho, C. J., and Vajda, S. (2002). Protein-protein association kinetics and protein docking. *Curr Opin Struct Biol* 12, 36-40.
- Charifson, P. S., Corkery, J. J., Murcko, M. A., and Walters, W. P. (1999). Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* 42, 5100-5109.
- Clarage, J. B., Romo, T., Andrews, B. K., Pettitt, B. M., and Phillips, G. N., Jr. (1995). A sampling problem in molecular dynamics simulations of macromolecules. *Proc Natl Acad Sci U S A* 92, 3288-3292.
- Clore, G. M. (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 U S A* 97, 9021-9025.
- Clore, G. M., and Schwieters, C. D. (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.
- Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221, 709-713.

- Dobrodumov, A., and Gronenborn, A. M. (2003). Filtering and selection of structural models: combining docking and NMR. *Proteins* 53, 18-32.
- Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 125, 1731-1737.
- Dominguez, C., Bonvin, A. M. J. J., Winkler, G. S., van Schaik, F. M. A., Timmers, H. T. M., and Boelens, R. (2004). Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis and docking approaches. *Structure* 12, 633-644.
- Fahmy, A., and Wagner, G. (2002). TreeDock:[?] A Tool for Protein Docking Based on Minimizing van der Waals Energies. *J Am Chem Soc* 124, 1241-1250.
- Gabb, H. A., Jackson, R. M., and Sternberg, M. J. (1997). Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* 272, 106-120.
- Gardiner, E. J., Willett, P., and Artymiuk, P. J. (2001). Protein docking using a genetic algorithm. *Proteins* 44, 44-56.
- Garrett, D. S., Seok, Y. J., Peterkofsky, A., Gronenborn, A. M., and Clore, G. M. (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.
- Halperin, I., Ma, B., Wolfson, H., and Nussinov, R. (2002). Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* 47, 409-443.
- Hoffmann, D., Kramer, B., Washio, T., Steinmetzer, T., Rarey, M., and Lengauer, T. (1999). Two-stage method for protein-ligand docking. *J Med Chem* 42, 4422-4433.
- Hus, J. C., Marion, D., and Blackledge, M. (2001). Determination of protein backbone structure using only residual dipolar couplings. *J Am Chem Soc* 123, 1541-1542.
- Jackson, R. M., Gabb, H. A., and Sternberg, M. J. (1998). Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J Mol Biol* 276, 265-285.
- Janin, J., Henrick, K., Moulton, J., Eyck, L. T., Sternberg, M. J., Vajda, S., Vakser, I., and Wodak, S. J. (2003). CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* 52, 2-9.
- Jiang, F., and Kim, S. H. (1991). "Soft docking": matching of molecular surface cubes. *J Mol Biol* 219, 79-102.
- Kay, L. E., Torchia, D. A., and Bax, A. (1989). Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28, 8972-8979.
- Knegtel, R. M., Antoon, J., Rullmann, C., Boelens, R., and Kaptein, R. (1994a). MONTY: a Monte Carlo approach to protein-DNA recognition. *J Mol Biol* 235, 318-324.
- Knegtel, R. M., Boelens, R., and Kaptein, R. (1994b). Monte Carlo docking of protein-DNA complexes: incorporation of DNA flexibility and experimental data. *Protein Eng* 7, 761-767.
- Knegtel, R. M. A., Kuntz, I. D., and Oshiro, C. M. (1997). Molecular docking to ensembles of protein structures. *J Mol Biol* 266, 424-440.
- Kohlbacher, O., Burchardt, A., Moll, A., Hildebrandt, A., Bayer, P., and Lenhof, H. P. (2001). Structure prediction of protein complexes by an NMR-based protein docking algorithm. *J Biomol NMR* 20, 15-21.
- Lee, W., Revington, M. J., Arrowsmith, C., and Kay, L. E. (1994). A pulsed field gradient isotope-filtered 3D ¹³C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. *FEBS Lett* 350, 87-90.

- Legrand, S., and Merz, K. (1992). The Application of Genetic Algorithm to Conformational Search. *Faseb J* 6, A132-A132.
- Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001). Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol* 339, 71-90.
- Lo Conte, L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J Mol Biol* 285, 2177-2198.
- McCoy, M. A., and Wyss, D. F. (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.
- Mendez, R., Leplae, R., De Maria, L., and Wodak, S. J. (2003). Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* 52, 51-67.
- Moont, G., Smith, G. R., and Stenberg, M. J. E. (March 2001). 3D-DOCK: incorporating FTDOCK (version 2.0), RPScore, and Multidock. <http://www.bmmicnetuk/docking/>.
- Morelli, X., Dolla, A., Czjzek, M., Palma, P. N., Blasco, F., Krippahl, L., Moura, J. J., and Guerlesquin, F. (2000). Heteronuclear NMR and soft docking: an experimental approach for a structural model of the cytochrome c553-ferredoxin complex. *Biochemistry* 39, 2530-2537.
- Morelli, X. J., Palma, P. N., Guerlesquin, F., and Rigby, A. C. (2001). A novel approach for assessing macromolecular complexes combining soft-docking calculations with NMR data. *Protein Sci* 10, 2131-2137.
- Nicholls, A., Bharadwaj, R., and Honig, B. (1993). Grasp - Graphical Representation and Analysis of Surface-Properties. *Biophys J* 64, A166-A166.
- Norel, R., Lin, S. L., Wolfson, H. J., and Nussinov, R. (1994). Shape complementarity at protein-protein interfaces. *Biopolymers* 34, 933-940.
- Norel, R., Petrey, D., Wolfson, H. J., and Nussinov, R. (1999). Examination of shape complementarity in docking of unbound proteins. *Proteins* 36, 307-317.
- Ogura, K., Terasawa, H., and Inagaki, F. (1996). An improved double-tuned and isotope-filtered pulse scheme based on a pulsed field gradient and a wide-band inversion shaped pulse. *J Biomol NMR* 8, 492-498.
- Otting, G., and Wüthrich, K. (1989). Extended Heteronuclear Editing of 2d H-1-Nmr Spectra of Isotope-Labeled Proteins, Using the X(Omega-1, Omega-2) Double Half Filter. *J Mag Res* 85, 586-594.
- Palma, P. N., Krippahl, L., Wampler, J. E., and Moura, J. J. (2000). BiGGER: a new (soft) docking algorithm for predicting protein interactions. *Proteins* 39, 372-384.
- Payne, A. W., and Glen, R. C. (1993). Molecular recognition using a binary genetic search algorithm. *J Mol Graph* 11, 74-91, 121-123.
- Peng, J. W., and Wagner, G. (1994). Investigation of Protein Motions Via Relaxation Measurements. In *Nuclear Magnetic Resonance, Pt C* (San Diego, ACADEMIC PRESS INC), pp. 563-596.
- Ritchie, D. W., and Kemp, G. J. (2000). Protein docking using spherical polar Fourier correlations. *Proteins* 39, 178-194.
- Smith, G. R., and Stenberg, M. J. (2002). Prediction of protein-protein interactions by docking methods. *Curr Opin Struct Biol* 12, 28-35.
- Terp, G. E., Johansen, B. N., Christensen, I. T., and Jorgensen, F. S. (2001). A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring

- (MultiScore) of protein-ligand binding affinities. *J Med Chem* *44*, 2333-2343.
- Tjandra, N., and Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* *278*, 1111-1114.
- Tjandra, N., Garrett, D. S., Gronenborn, A. M., Bax, A., and Clore, G. M. (1997). Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nat Struct Biol* *4*, 443-449.
- Tolman, J. R., Flanagan, J. M., Kennedy, M. A., and Prestegard, J. H. (1995). Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc Natl Acad Sci U S A* *92*, 9279-9283.
- Totrov, M., and Abagyan, R. (1997). Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins Suppl* *1*, 215-220.
- Ubbink, M., Ejdeback, M., Karlsson, B. G., and Bendall, D. S. (1998). The structure of the complex of plastocyanin and cytochrome *f*, determined by paramagnetic NMR and restrained rigid-body molecular dynamics. *Structure* *6*, 323-335.
- Venclovas, C., Zemla, A., Fidelis, K., and Moutl, J. (2003). Assessment of progress over the CASP experiments. *Proteins* *53 Suppl* *6*, 585-595.
- Vieth, M., Hirst, J. D., Dominy, B. N., Daigler, H., and Brooks, C. L. (1998). Assessing search strategies for flexible docking. *J Comput Chem* *19*, 1623-1631.
- Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J Mol Graph* *8*, 52-56, 29.
- Weber, D. J., Gittis, A. G., Mullen, G. P., Abeygunawardana, C., Lattman, E. E., and Mildvan, A. S. (1992). NMR docking of a substrate into the X-ray structure of staphylococcal nuclease. *Proteins* *13*, 275-287.
- Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids* (New York, John Wiley & Sons).
- Xu, X. P., and Case, D. A. (2001). Automated prediction of ¹⁵N, ¹³C_α, ¹³C_β and ¹³C' chemical shifts in proteins using a density functional database. *J Biomol NMR* *21*, 321-333.
- Zuiderweg, E. R. (2002). Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* *41*, 1-7.
- Zwahlen, C., Legault, P., Vincent, S. J. F., Greenblatt, J., Konrat, R., and Kay, L. E. (1997). Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *J Am Chem Soc* *119*, 6711-6721.



HADDOCK: a protein-protein docking approach based on biochemical or biophysical data

Cyril Dominguez, Rolf Boelens, and Alexandre M. J. J. Bonvin

J. Am. Chem. Soc., 125, 1731-1737 (2003)

Reproduced with permission of the American Chemical Society

Abstract

The structure determination of protein–protein complexes is a rather tedious and lengthy process, by both NMR and X-ray crystallography. Several methods based on docking to study protein complexes have also been well developed over the past few years. Most of these approaches are not driven by experimental data but are based on a combination of energetics and shape complementarity. Here we present an approach called HADDOCK (High Ambiguity Driven protein–protein DOCKing) that makes use of biochemical and/or biophysical interaction data such as chemical shift perturbation data resulting from NMR titration experiments or mutagenesis data. This information is introduced as Ambiguous Interaction Restraints (AIRs) to drive the docking process. An AIR is defined as an ambiguous distance between all residues shown to be involved in the interaction. The accuracy of our approach is demonstrated with three molecular complexes. For two of these complexes, for which both the complex and the free protein structures have been solved, NMR titration data were available. Mutagenesis data were used in the last example. In all cases, the best structures generated by HADDOCK, that is, the structures with the lowest intermolecular energies, were the closest to the published structure of the respective complexes (within 2.0 Å backbone RMSD).

Introduction

For a better understanding of the biological function of a protein, knowledge of its three-dimensional structure is crucial. Solving protein structures is mainly achieved by two different methods: X-ray crystallography and Nuclear Magnetic Resonance (NMR). From the statistics of the Protein Data Bank (PDB) (Berman et al., 2000) (<http://www.rcsb.org/pdb/>), approximately 13500 X-ray structures and 2225 NMR structures have been solved and deposited at this date. Most of the proteins achieve their function by interacting with other proteins and forming an active complex. Although many methods are available to study protein complexes at different levels (two-hybrid screening, fluorescence studies, resonance energy transfer, etc), only few of these techniques provide high-resolution information at an atomic level. X-ray and NMR encounter difficulties in dealing with structures of complexes. Indeed, by X-ray, the dynamic of the complex formation makes the crystallization difficult, while the size limitation in NMR is a major problem when considering high molecular weight complexes. The traditional NMR approach to solving protein–protein complexes requires the collection of intermolecular nuclear Overhauser effect (NOE) distances, which is typically a lengthy and difficult process. In addition, intermolecular NOEs often involve side chain protons, requiring thus a rather complete assignment of all NMR signals. Because of these limitations, the number of protein–protein complexes solved and deposited in the PDB is rather low (643 by X-ray and 84 by NMR) compared with the number of free form structures. NMR, however, is very powerful in mapping protein–protein interfaces by titration experiments (reviewed in Zuiderweg, 2002). Such experiments, which can be

performed at the stage of backbone assignment already, easily allow us to identify amino acids involved in the complex formation but do not provide any information about the orientation of one protein with respect to its counterpart. Because of that, this information has rarely been directly used as a structural restraint in a structure calculation process. Next to these experimental approaches, theoretical methods to study protein complexes at a structural level based on docking are now emerging that have been well developed during the past few years. There are now a number of programs performing “ab initio” protein-protein docking (for review, see references Camacho and Vajda, 2002; Smith and Sternberg, 2002). Most of these programs use the same approach: one protein is fixed in space and the second one is rotated and translated around the first one. For each new configuration, a score is calculated on the basis of various terms such as surface complementarities, electrostatic interactions, van der Waals repulsion, and so forth. The drawback of these methods is that the search through the entire conformational space of the complex geometry makes the calculation heavy, rarely resulting in a unique solution. Recently, NMR data have been used in combination with docking methods in different ways to generate protein-protein complexes. Diamagnetic chemical shift changes and intermolecular pseudocontact shifts have been combined with restrained rigid-body molecular dynamics to solve the structure of the paramagnetic plastocyanin-cytochrome *f* complex (Ubbink et al., 1998). Intermolecular NOEs and residual dipolar couplings (RDCs) have been combined to solve the structure of the EIN-HP α complex (Clare, 2000). Intermolecular NOEs can very accurately define the interface. Their collection, however, is generally a tedious process. In addition, RDC data can be very useful to determine the relative orientation of the two proteins. Morelli et al. used the program BIGGER (Palma et al., 2000), which makes use of an NMR filter on the basis of chemical shift perturbation data (Morelli et al., 2001). This approach allows the use of NMR titration data to rank the possible solutions, but the docking is not directly driven by these data. Recently, Fahmy et al. developed a new docking program, TreeDock, where the docking is oriented on the basis of anchors points which can be in principle derived from NMR chemical shift perturbation or mutagenesis data (Fahmy and Wagner, 2002). This program performs a rigid body docking and the solutions are ranked in function of their Lennard-Jones potentials. McCoy et al. used chemical shift perturbation data in combination with RDCs to develop a new docking approach (McCoy and Wyss, 2002). In that case, the RDC data are first introduced to orient the complex, and then the solutions are optimized by back calculating chemical shift perturbation with the SHIFTS software (Xu and Case, 2001) and comparing them with the experimental data. During complex formation, usually, some structural rearrangements occur. By NMR titration, it is possible to check such rearrangements for backbone atoms, but no information is available on the side chain rearrangements that occur frequently at the interface, especially in the case of hydrophobic interfaces. It is therefore important, when docking two proteins, to consider the best orientation of their side chains leading to the minimum energy and the best side chain contacts. For this, the side chains at the interface should be free to adapt their conformation.

Here, we present a new high ambiguity driven docking approach (HADDOCK) that

makes use of biochemical and/or biophysical interaction data such as, for example, chemical shift perturbation data obtained from NMR titration experiments or mutagenesis data. The information on the interacting residues is introduced as Ambiguous Interaction Restraints (AIR) to drive the docking. After calculation, the structures are ranked according to their intermolecular energy, that is, sum of electrostatic, van der Waals, and AIR energy terms. We should note that ambiguous distance restraints have first been introduced to solve symmetric dimer structures by NMR (Nilges, 1993) and are now commonly used in protein structure determination and automated NOE assignment methods (Nilges and Donoghue, 1998). We demonstrate the usefulness of the AIRs and the accuracy of our docking approach for three different molecular complexes: the N-terminus domain of Enzyme I (EIN) in complex with the histidine-containing phosphocarrier protein (HPr), the Enzyme IIA^{glucose} (EIIA) in complex with HPr and the HIV protein gp120 in complex with the protein CD4. The structures of the first two complexes have been solved by NMR, (Garrett et al., 1999; Wang et al., 2000) and their respective free forms are available from X-ray and/or NMR (Jia et al., 1993; Liao et al., 1996; van Nuland et al., 1994; Worthylake et al., 1991). The NMR titration data of each protein upon complex formation to its partner are available (Chen et al., 1993; Garrett et al., 1997; van Nuland et al., 1995). For the gp120-CD4 complex, however, only the X-ray structures (Kwong et al., 2000; Kwong et al., 1998) of the individual partners of a complex were used. Instead of NMR titration data, mutagenesis data (Moebius et al., 1992; Olshevsky et al., 1990) were used to define ambiguous interaction restraints. In all three cases, starting from the complex or the free state structures, we found that the best solutions generated by HADDOCK, that is, the structures with the lowest intermolecular energy term, were those that are the closest in terms of backbone root-mean-square deviations at the interface (iRMSD) (between 0.8 and 2 Å) to the published structure of the respective complexes.

Results

Ambiguous Interaction Restraints (AIR)

The Ambiguous Interaction Restraints are derived from any kind of experimental information available concerning residues that are involved in the intermolecular interaction. We distinguish here between “active” and “passive” residues. In the case of NMR titration data, the active residues correspond to all residues showing a significant chemical shift perturbation upon complex formation as well as a high solvent accessibility in the free form protein (>50% relative accessibility as calculated with NACCESS (Hubbard and Thornton, 1993)). The threshold to define significant chemical shift perturbations will differ for each protein complex under study and need some optimization by the user. In our examples, we used as starting point the residues that the authors of the original papers (Chen et al., 1993; Garrett et al., 1997; van Nuland et al., 1995) defined as significantly perturbed in the complex. These perturbed residues that do not satisfy the high solvent accessibility criterion

should be subsequently removed from the active residue list. In the case of mutagenesis data, the active residues are those that have been shown by mutations to alleviate complex formation and are also solvent exposed. The passive residues correspond to the residues that show a less significant chemical shift perturbation and/or that are surface neighbors of the active residues and have a high solvent accessibility (>50%). An AIR is defined as an ambiguous intermolecular distance (d_{iAB}) with a maximum value of 3 Å between any atom m of an active residue i of protein A (m_{iA}) and any atom n of both active and passive residues k (N_{res} in total) of protein B (n_{kB}) (and inversely for protein B). The effective distance d_{iAB}^{eff} for each restraint is calculated using the equation:

$$d_{iAB}^{eff} = \left(\sum_{m_{iA}=1}^{N_{atoms}} \sum_{k=1}^{N_{resB}} \sum_{n_{kB}=1}^{N_{atoms}} \frac{1}{d_{m_{iA}n_{kB}}^6} \right)^{-\frac{1}{6}}$$

where N_{atoms} indicates all atoms of a given residue and N_{res} the sum of active and passive residues for a given protein. In this way, the passive residues do not have direct AIRs to the partner protein but can satisfy the partner protein active restraints. A $1/r^6$ sum averaging is used, not by analogy to NOE restraints, but because this mimics the attractive part of a Lennard-Jones potential and ensures that the AIRs are satisfied as soon as any two atoms of the two proteins are in contact. The 3 Å limit represents a compromise between hydrogen-hydrogen and heavy atom-heavy atom minimum van der Waals distances. The use of ambiguous interaction restraints allows HADDOCK to search through all the possible configurations around the interacting site defined by the biochemical and/or biophysical data such as NMR chemical shift perturbation data or mutagenesis data and to find the most favorable pair of interacting amino acids among the active and passive residues.

Docking Protocol

Our HADDOCK (High Ambiguity Driven protein-protein DOCKing) has been implemented in CNS (Brunger et al., 1998) for structure calculations and makes use of python scripts derived from ARIA (Linge et al., 2001) for automation (see Material and Methods). The docking protocol, which requires the PDB files of the free proteins and ambiguous interaction restraints, consists of three stages: i) randomization of orientations and rigid body energy minimization (EM), ii) semi rigid simulated annealing in torsion angle space (TAD-SA), iii) final refinement in Cartesian space with explicit solvent.

The three stages are detailed in the Material and Methods section. During the TAD simulated annealing and the water refinement, the amino acids at the interface (side chains and backbone) are allowed to move to optimize the interface packing. The interface amino acids allowed to move are defined by the active and passive amino acids used in the AIRs ± 2 sequential amino acids. Although no real significant structural changes occur during the water refinement stage, it is useful for the improvement of the energetics of the interface. This is important for a proper scoring of the resulting conformations. The final structures

are clustered using the pairwise backbone RMSD at the interface and analyzed according to their average interaction energies (sum of E_{elec} , E_{vdw} , E_{AIR}) and their average buried surface area. The entire docking procedure is performed automatically by HADDOCK and is followed by the cluster analysis (for more details, see Material and Methods). For the EIN-HPr complex (247 and 85 amino acids, 25 AIRs requiring 105000 distance evaluations), the entire run required 2 days on 10 1.3 GHz AMD processors. The three docking stages required 10 s, 1.5 h and 1 h per structure for the rigid body minimization, the semirigid TAD-SA and the final water refinement, respectively.

Validation of the HADDOCK Approach

HADDOCK was tested for three protein-protein complexes using chemical shift perturbation data in two cases and mutagenesis data in the third to define the ambiguous interaction restraints. As a first test, we performed the docking with ambiguous interaction restraints on the EIN-HPr complex (Garrett et al., 1999) starting from the structure of the complex. The coordinates of the two proteins in the structure of the complex were separated into two distinct pdb files. Although the structures of the two proteins and in particular of their interface were already in the geometry of the complex, the side chains and backbone atoms at the interface were still allowed to move during the TAD simulated annealing and the water refinement process. On the basis of the NMR titration data (Garrett et al., 1997; van Nuland et al., 1995), 24 amino acids of EIN and 19 amino acids of HPr showing significant chemical shift perturbation were first identified. The solvent accessibility of these amino acids was calculated, and only those that are exposed at the surface of the protein were further selected for the active ambiguous interaction restraints. At the end, 16 amino acids of EIN (E67, E68, K69, A71, I72, D82, E83, E84, G110, Q111, S113, A114, E116, E117, L118 and Y122) and 9 amino acids of HPr (H15, T16, R17, Q21, K24, K49, Q51, T52, and G54) were used as active AIRs. By displaying these amino acids on the free form structures, we defined five passive amino acids for EIN (M78, L79, L115, L123 and R126) and three for HPr (A20, L47 and F48). The interface residues that were allowed to move during the TAD simulated annealing and the water refinement process consisted of residues 65 to 74, 76 to 86 and 108 to 128 for EIN and 13 to 26 and 45 to 56 for HPr. Figure 1 (circles) shows the intermolecular energy as a function of the iRMSD (backbone RMSD at the interface) from the target, that is, the NMR structure, for the 200 calculated structures after water refinement. Five clusters were obtained. Their average intermolecular energies are, respectively, -868, -698, -465, -270, and -388 kcal mol⁻¹ and the average iRMSDs from the target are 1.4, 2.7, 8.5, 9.0, and 9.5 Å. For reference, the published NMR structure that has however not been optimized within our chosen force field and parameters has an intermolecular energy of -370 kcal mol⁻¹. Cluster 1 has the lowest intermolecular energy as well as the lowest iRMSD from the target. This result demonstrates a nice correlation between the intermolecular energy of our solutions and the iRMSD between these solutions and the target. The best solution of Cluster 1 (the lowest in energy) has an intermolecular energy of -961 kcal mol⁻¹ and an iRMSD of 1.45 Å (the backbone RMSD on both proteins

is 1.05 Å) from the reference structure (Figure 2A).

Next, HADDOCK was run, starting from the protein structures in the free form (Jia et al., 1993; Liao et al., 1996). The backbone iRMSD between the free and bound form of EIN and HPr are 0.95 and 0.55 Å, respectively. The resulting intermolecular energies as a function of the iRMSD from the target for the 200 calculated structures after water refinement are shown in Figure 1 (triangles). After analysis, 13 clusters were obtained with average energies between -637 and -275 kcal mol⁻¹ and average iRMSDs from the target between 1.80 and 9.85 Å. Again, in this case, the lowest intermolecular energy cluster corresponds to the lowest iRMSD from the target. The best solution of this cluster has an intermolecular energy of -658 kcal mol⁻¹ and an iRMSD from the target of 1.70 Å (the backbone RMSD on both proteins is 2.75 Å) (Figure 2C). These results demonstrate that HADDOCK could generate the correct docking solution starting from the free form protein structures and that, again, the lowest intermolecular energy cluster is the closest one to the published NMR structure. Among all protein-protein complexes available in the PDB, the average buried interface area is 1600 ± 400 Å² (Lo Conte et al., 1999). In our case, the best solutions have a buried interface area of 2064 Å² when starting from the complex form and 1798 Å² when starting from the free form proteins, while the buried surface area of the NMR structure of the complex is 1996 Å².

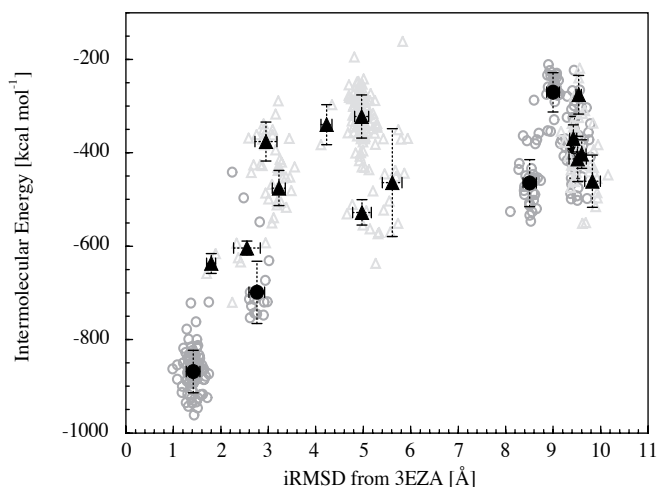


Figure 1: Intermolecular energies versus iRMSDs for the EIN-HPr complex. The energies are calculated as the sum of $E_{\text{elec}} + E_{\text{vdw}} + E_{\text{AIR}}$ after water refinement. iRMSD corresponds to the backbone RMSD at the interface from the pdb structure (3EZA). (Open circles) Single conformations (200) and (Filled circles) cluster averages when starting from the complex conformation. (Open triangle) Single conformations (200) and (Filled triangle) cluster averages when starting from the free form structures.

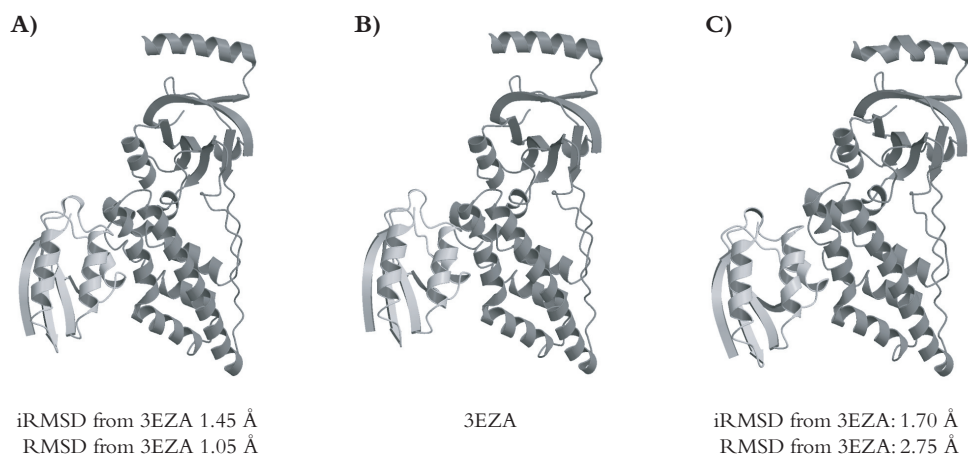


Figure 2: Comparison of the EIN–HPr solutions generated by HADDOCK with the reference structure. **A)** Best solution of lowest energy cluster when starting from the structures of the complex. **B)** Reference structure (PDB:3EZA). **C)** Best solution of lowest energy cluster when starting from the free form structures. iRMSD corresponds to the backbone RMSD at the interface from the reference structure. These figures have been generated with the programs Molscript (Kraulis, 1991) and Raster3D (Merritt and Murphy, 1994). HPr is represented in light grey.

As a second test, the structure of the EIIA–HPr complex (Wang et al., 2000) starting from the free form protein structures (Jia et al., 1993; Worthylake et al., 1991) was calculated with HADDOCK. The backbone iRMSD between the free and bound form of EIIA and HPr are 0.35 and 0.05 Å, respectively. The intermolecular energy of the complex is $-207 \text{ kcal mol}^{-1}$ and its buried surface area is 1434 Å^2 . We defined the AIRs as previously described, selecting 11 active (D38, V40, I45, V46, K69, F71, S78, E80, D94, V96 and S141) and 4 passive amino acids (V39, G68, E72 and E86) for EIIA and 9 active (H15, T16, R17, A20, F48, Q51, T52, G54 and T56) and 1 passive (N12) amino acid for HPr. The flexible interface consisted of amino acids 36 to 48, 66 to 82, 84 to 88, 92 to 98, and 139 to 143 for EIIA and 10 to 22, and 46 to 58 for HPr. The resulting intermolecular energies as a function of the iRMSD from target for the 200 calculated structures after water refinement are shown in Figure 3. Clusters (27) were obtained with average intermolecular energies between -453 and $-69 \text{ kcal mol}^{-1}$ and average iRMSDs from the published structure between 2.0 and 9.9 Å. Again, the lowest energy cluster is the one that is the closest to the reference structure. Its best solution has an intermolecular energy of $-493 \text{ kcal mol}^{-1}$, an iRMSD from the published structure of 2.10 Å (the backbone RMSD on both proteins is 2.30 Å) and a buried surface area of 1404 Å^2 (Figure 4).

We finally tested the feasibility of using data from mutagenesis studies to define ambiguous interaction restraints to drive the docking process. For this, docking was performed on the gp120–CD4 complex (Kwong et al., 2000). The intermolecular energy of the complex

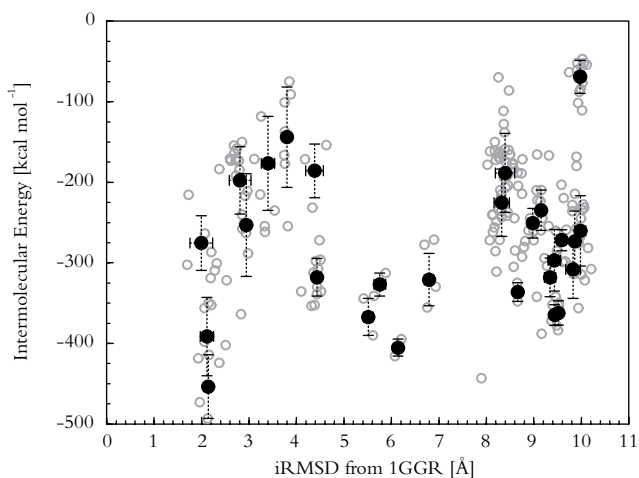


Figure 3: Intermolecular energies versus iRMSDs for the EIHA-HPPr complex. Energies and iRMSDs as defined in Figure 1. The pdb code of the reference structure is 1GGR. (Open circle) Single conformations (200) and (Filled circles) cluster averages when starting from the free form structures.

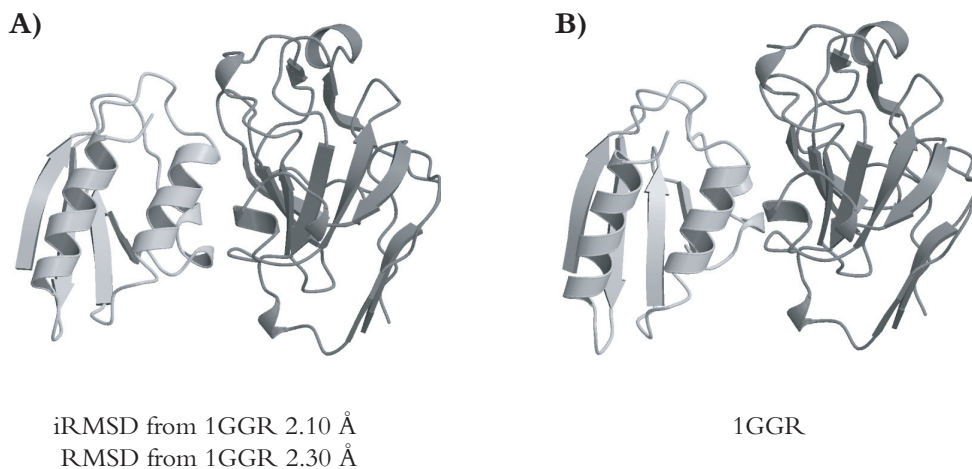


Figure 4: Comparison of the EIHA-HPPr best HADDOCK solution with the reference structure. **A)** Best solution of lowest energy cluster. **B)** Reference structure (PDB:1GGR). iRMSD as defined in Figure 2. HPr is represented in light grey.

solved by crystallography is $-283 \text{ kcal mol}^{-1}$ and the buried surface area is 1990 \AA^2 . To speed-up the calculation, the C-terminus domain of CD4 that does not interact with gp120 was removed and only residues 90 to 492 of gp120 and residues 1 to 97 of CD4 were used. The separated forms of the complex were used as the starting point. Mutagenesis data have revealed that residues D368, E370, W427, and D457 of gp120 and residues K29, K35, F43, L44, K46, G47, and R59 of CD4 were important for the binding (Moebius et al., 1992; Olshevsky et al., 1990). These amino acids have been used as active residues in the AIRs. In addition, 19 amino acids for gp120 (I109, N280, A281, K362, S365, G367, I371, N425, K429, V430, T455, G459, I467, R469, G471, G472, G473, D474 and R476) and 10 amino acids for CD4 (H27, Q33, I34, Q40, S42, T45, P48, N52, D53 and D56) were selected as passive residues. The flexible interface consisted of amino acids 107 to 111, 278 to 283, 360 to 373, 423 to 432, 453 to 461, and 465 to 478 for gp120 and 26 to 62 for CD4. The resulting intermolecular energies as a function of the iRMSD from target for the 200 calculated structures after water refinement are shown in Figure 5. Clusters (8) were obtained with average intermolecular energies between -407 and $-139 \text{ kcal mol}^{-1}$ and average backbone iRMSDs from the target between 0.9 and 11.5 \AA . Again, a nice correlation between the intermolecular energy and the iRMSD from the target for the clusters is observed. The best solution from the lowest energy cluster has an intermolecular energy of $-445 \text{ kcal mol}^{-1}$, an iRMSD from the published structure of 0.80 \AA (the backbone RMSD on both proteins is 0.80 \AA) and a buried surface area of 2148 \AA^2 (Figure 6).

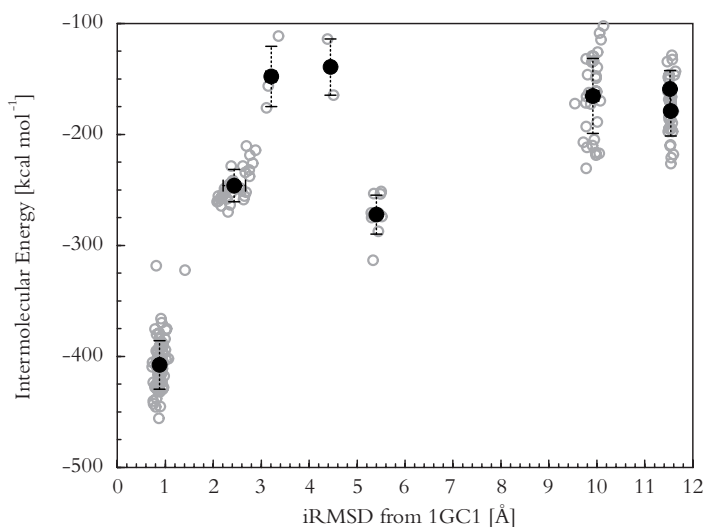


Figure 5: Intermolecular energies versus iRMSDs for the gp120-CD4 complex. Energies and iRMSDs as defined in Figure 1. The pdb code of the reference structure is 1GC1. (Open circles) Single conformations (200) and (Filled circles) cluster averages when starting from the complex conformation.

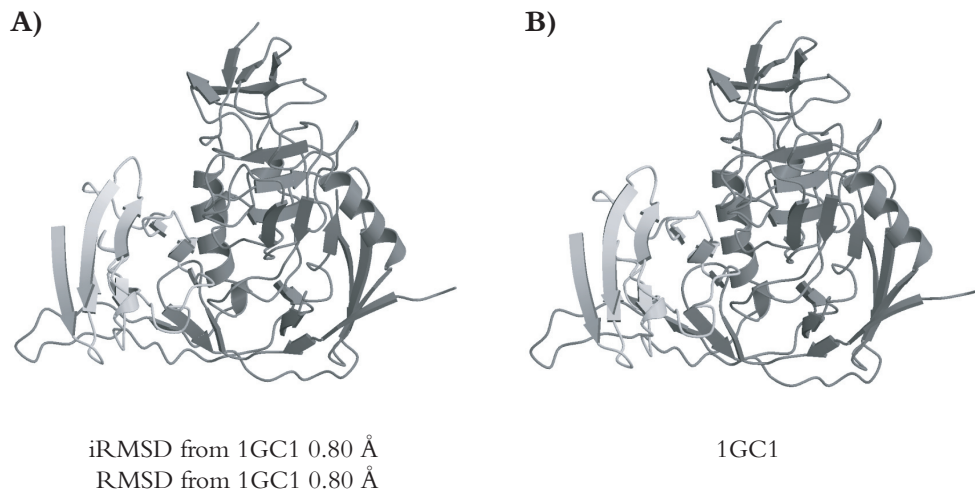


Figure 6: Comparison of the gp120–CD4 best HADDOCK solution with the reference structure. **A)** Best solution of lowest energy cluster. **B)** Reference structure (PDB:1GC1). iRMSD as defined in Figure 2. CD4 is represented in light grey.

These results nicely demonstrate that biochemical interaction data such as mutagenesis data can also be used to define highly ambiguous restraints to drive the docking with HADDOCK.

Discussion

We have developed an approach, HADDOCK, that allows rapid and accurate docking of protein complexes based on the use of biochemical or biophysical information. This information, which is introduced as ambiguous interaction restraints, is sufficient to drive the docking process. It is important to note that to reduce considerably the ambiguity, information about the interfaces of both proteins is needed. On the basis of the intermolecular energy, the lowest energy clusters generated by HADDOCK were in all cases the closest to the published structure. The fact that the side chains at the interface are allowed to move increase the accuracy of our scoring compared with classical rigid body docking. Indeed, simulated annealing and water refinement do not improve much the iRMSD from the target, but by allowing the side chains to reorient and adopt better conformations, a better scoring of the solutions (a good correlation between the intermolecular energy and the iRMSD from the target) is obtained. The AIR restraints that we have used in the three examples contain, in principle, no information on the relative orientation of the two partners in the complex. Indeed, 180° rotated solutions are obtained that have quite low intermolecular

energies (see for example Figure 3). The discrimination between orientations must therefore come mainly from the van der Waals and electrostatic energy terms. This is made possible because of some degree of asymmetry at the interface both in shape complementarities and in the distribution of hydrophobic and hydrophilic residues. One should thus realize that a correct scoring of solutions will depend on the nature of the interface and that, without additional experimental information, the scoring might not be as effective in the case of complexes lacking some kind of asymmetry in their interface.

The power of our approach has been demonstrated using chemical shift perturbation data and mutagenesis data, but any kind of data that provides information on the interaction interface could in principle be used to drive the docking and to improve the validity of the solutions. This could be additional NMR restraint such as intermolecular NOEs, RDCs, but also, other types of biochemical or biophysical interaction data could be considered. In this work, the ambiguous interaction restraints (AIRs) were defined with a conservative fixed distance of 3.0 Å. This value could be optimized by differentiating the strength of restraints, depending on a scaling of the distance as a function of the chemical shift perturbation in hertz and/or the type of amino acid. Though this may provide a more accurate and precise scoring, our results show that meaningful structures are already produced with simple and conservative restraints, demonstrating the robustness of our approach. It is also clear in the case of chemical shift perturbation data that better experimental data can be obtained. By using ^{15}N - ^{13}C double-labeled proteins, side chain information can be collected that will allow a more precise definition of the side chains atoms that are implicated in the interaction. This information could be important to refine the ambiguous interaction restraints (AIRs) and thus the accuracy of HADDOCK.

Material and Methods

Structural Coordinates

The coordinates of all proteins in free and bound form were obtained from the protein data bank (PDB) (Berman et al., 2000). The accession number for the EIN-HPr complex (Garrett et al., 1999), the free EIN (Liao et al., 1996) and the free HPr (Jia et al., 1993) are respectively 3EZA, 1ZYM and 1POH. The accession number of the EIIA-HPr complex (Wang et al., 2000) and the free form of EIIA (Worthylake et al., 1991) are, respectively, 1GGR and 1F3G. The accession number of the gp120-CD4 complex (Kwong et al., 2000) is 1GC1.

Docking Protocol

Our HADDOCK (High Ambiguity Driven DOCKing) approach consists of a collection of python scripts derived from ARIA (Linge et al., 2001) and makes use of CNS (Brunger et al., 1998) for structure calculation. The python scripts take care of setting up the system from the PDB files of the free proteins, of carrying and monitoring the structure calculations, and of sorting and analyzing the docking solutions. Inter- and intramolecular

energies are evaluated using full electrostatic and van der Waals energy terms with an 8.5 Å distance cut-off using the OPLS nonbonded parameters (Jorgensen and Tirado-rives, 1998) from a modified version of the parallhdg5.2.pro parameter file (Linge and Nilges, 1999) (Marc Williams, University College London, personal communication). The docking protocol consists of three stages: i) randomization of orientations and rigid body energy minimization (EM), ii) semi rigid simulated annealing in torsion angle space (TAD-SA), iii) final refinement in Cartesian space with explicit solvent.

In the randomization stage, the two partner proteins are positioned at 150 Å from each other in space and each protein is randomly rotated around its center of mass. Rigid body EM is then performed: the first four cycles of orientational optimization are performed in which each protein in turn is allowed to rotate to minimize the intermolecular energy function. Then both translations and rotations are allowed, and the two proteins are docked by rigid body EM. Typically 1000 complex conformations are calculated at this stage. The best 200 solutions in terms of intermolecular energies are then refined. The second stage consists of three simulated annealing refinements. In the first simulated annealing (1000 steps from 2000 to 50 K with 8 fs time steps), the two proteins are considered as rigid bodies and their respective orientation is optimized. In the second simulated annealing (4000 steps from 2000 to 50 K with 4 fs time steps), the side chains at the interface are allowed to move. In the third simulated annealing (1000 steps from 500 to 50 K with 2 fs time steps), both side chains and backbone at the interface are allowed to move to allow for some conformational rearrangements. The resulting structures are then subjected to 200 steps of steepest descent EM. The final stage consists of a gentle refinement in an 8 Å shell of TIP3P water molecules (Jorgensen et al., 1992). A 2 fs time step is used for the integration of the equation of motions. The system is first heated to 300 K (500 steps at 100, 200, and 300 K) with position restraints ($k_{\text{pos}} = 5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) on all atoms except for the flexible side chains at the interface. MD steps (5000) are then performed at 300 K with position restraints only on noninterface heavy atoms ($k_{\text{pos}} = 1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$). During the final cooling stage (1000 MD steps at 300, 200, and 100 K), the position restraints are limited to backbone atoms outside the interface. The final structures are clustered using the pairwise backbone RMSD at the interface. A cluster is defined as an ensemble of at least two conformations displaying an iRMSD (backbone RMSD at the interface) smaller than 1.0 Å. The resulting clusters are analyzed and ranked according to their average interaction energies (sum of E_{elec} , E_{vdw} , E_{ACS}) and their average buried surface area.

The HADDOCK package will be made available upon request. In a similar manner as the ARIA program (Nilges et al., 1997), HADDOCK can be set up via a Web browser interface that makes it user-friendly. All the parameters that we used in our examples are set up as default parameters but can be modified by the user to possibly optimize the protocols for a particular problem.

Acknowledgment

This work was supported by a grant from the European community (5th Framework program NMRQUAL Contract Number QLG2-CT-2000-01313) and a “Jonge Chemici” grant from The Netherlands Organization for Scientific Research (NWO) to Dr. A. M. J. J. Bonvin. Financial support from the Center for Biomedical Genetics is also acknowledged. The authors are grateful to Shang–Te Hsu for useful discussions.

References

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res* 28, 235-242.
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54 (Pt 5), 905-921.
- Camacho, C. J., and Vajda, S. (2002). Protein-protein association kinetics and protein docking. *Curr Opin Struct Biol* 12, 36-40.
- Chen, Y., Reizer, J., Saier, M. H., Jr., Fairbrother, W. J., and Wright, P. E. (1993). Mapping of the binding interfaces of the proteins of the bacterial phosphotransferase system, HPr and IIAGlc. *Biochemistry* 32, 32-37.
- Clore, G. M. (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 U S A* 97, 9021-9025.
- Fahmy, A., and Wagner, G. (2002). TreeDock:[?] A Tool for Protein Docking Based on Minimizing van der Waals Energies. *J Am Chem Soc* 124, 1241-1250.
- Garrett, D. S., Seok, Y. J., Peterkofsky, A., Clore, G. M., and Gronenborn, A. M. (1997). Identification by NMR of the binding surface for the histidine-containing phosphocarrier protein HPr on the N-terminal domain of enzyme I of the Escherichia coli phosphotransferase system. *Biochemistry* 36, 4393-4398.
- Garrett, D. S., Seok, Y. J., Peterkofsky, A., Gronenborn, A. M., and Clore, G. M. (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.
- Hubbard, S. J., and Thornton, J. M. (1993). NACCESS. “NACCESS”, Department of biochemistry and molecular Biology, University College London.
- Jia, Z., Quail, J. W., Waygood, E. B., and Delbaere, L. T. (1993). The 2.0-Å resolution structure of Escherichia coli histidine-containing phosphocarrier protein HPr. A redetermination. *J Biol Chem* 268, 22490-22501.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1992). *J Chem Phys* 96, 926-935.
- Jorgensen, W. L., and Tirado-rives, J. (1998). The OPLS Potential functions for proteins. Energy minimizations for crystals of cyclin peptides and crambin. *J Am Chem Soc* 110, 1657-1666.

- Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24, 946-950.
- Kwong, P. D., Wyatt, R., Majeed, S., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (2000). Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure Fold Des* 8, 1329-1339.
- Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393, 648-659.
- Liao, D. I., Silvertown, E., Seok, Y. J., Lee, B. R., Peterkofsky, A., and Davies, D. R. (1996). The first step in sugar transport: crystal structure of the amino terminal domain of enzyme I of the E. coli PEP: sugar phosphotransferase system and a model of the phosphotransfer complex with HPr. *Structure* 4, 861-872.
- Linge, J. P., and Nilges, M. (1999). Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. *J Biomol NMR* 13, 51-59.
- Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001). Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol* 339, 71-90.
- Lo Conte, L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J Mol Biol* 285, 2177-2198.
- McCoy, M. A., and Wyss, D. F. (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.
- Merrit, E. A., and Murphy, M. E. P. (1994). Raster3D version 2.0: A program for photorealistic molecular graphics. *Acta Cryst D* 50, 869-873.
- Moebius, U., Clayton, L. K., Abraham, S., Harrison, S. C., and Reinherz, E. L. (1992). The human immunodeficiency virus gp120 binding site on CD4: delineation by quantitative equilibrium and kinetic binding studies of mutants in conjunction with a high-resolution CD4 atomic structure. *J Exp Med* 176, 507-517.
- Morelli, X. J., Palma, P. N., Guerlesquin, F., and Rigby, A. C. (2001). A novel approach for assessing macromolecular complexes combining soft-docking calculations with NMR data. *Protein Sci* 10, 2131-2137.
- Nilges, M. (1993). A calculation strategy for the structure determination of symmetric dimers by ¹H NMR. *Proteins* 17, 297-309.
- Nilges, M., and Donoghue, S. I. (1998). Ambiguous NOEs and automated NOE assignment. *Prog Nucl Mag Res Sp* 32, 107-139.
- Nilges, M., Macias, M. J., O'Donoghue, S. I., and Oschkinat, H. (1997). Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol* 269, 408-422.
- Olshevsky, U., Helseth, E., Furman, C., Li, J., Haseltine, W., and Sodroski, J. (1990). Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 receptor binding. *J Virol* 64, 5701-5707.
- Palma, P. N., Krippahl, L., Wampler, J. E., and Moura, J. J. (2000). BiGGER: a new (soft) docking algorithm for predicting protein interactions. *Proteins* 39, 372-384.
- Smith, G. R., and Sternberg, M. J. (2002). Prediction of protein-protein interactions by docking

methods. *Curr Opin Struct Biol* 12, 28-35.

Ubbink, M., Ejdeback, M., Karlsson, B. G., and Bendall, D. S. (1998). The structure of the complex of plastocyanin and cytochrome f, determined by paramagnetic NMR and restrained rigid-body molecular dynamics. *Structure* 6, 323-335.

van Nuland, N. A., Boelens, R., Scheek, R. M., and Robillard, G. T. (1995). High-resolution structure of the phosphorylated form of the histidine-containing phosphocarrier protein HPr from *Escherichia coli* determined by restrained molecular dynamics from NMR-NOE data. *J Mol Biol* 246, 180-193.

van Nuland, N. A., Hangyi, I. W., van Schaik, R. C., Berendsen, H. J., van Gunsteren, W. F., Scheek, R. M., and Robillard, G. T. (1994). The high-resolution structure of the histidine-containing phosphocarrier protein HPr from *Escherichia coli* determined by restrained molecular dynamics from nuclear magnetic resonance nuclear Overhauser effect data. *J Mol Biol* 237, 544-559.

Wang, G., Louis, J. M., Sondej, M., Seok, Y. J., Peterkofsky, A., and Clore, G. M. (2000). Solution structure of the phosphoryl transfer complex between the signal transducing proteins HPr and IIA(glucose) of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system. *Embo J* 19, 5635-5649.

Worthylake, D., Meadow, N. D., Roseman, S., Liao, D. I., Herzberg, O., and Remington, S. J. (1991). Three-dimensional structure of the *Escherichia coli* phosphocarrier protein III_{glc}. *Proc Natl Acad Sci U S A* 88, 10382-10386.

Xu, X. P., and Case, D. A. (2001). Automated prediction of ¹⁵N, ¹³C_{alpha}, ¹³C_{beta} and ¹³C' chemical shifts in proteins using a density functional database. *J Biomol NMR* 21, 321-333.

Zuiderweg, E. R. (2002). Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* 41, 1-7.



NMR solution structure of the E2 ubiquitin conjugating enzyme UbCH5B

Cyril Dominguez,* Klaartje Houben,* Frederik M. A. van Schaik, H. Th.
Marc Timmers, Alexandre M. J. J. Bonvin and Rolf Boelens

In preparation

* These authors contributed equally to this work

Abstract

The structure of the ubiquitin conjugating enzyme (E2) UbcH5B has been solved by a combination of homology modeling, diffusion anisotropy derived from NMR relaxation data and automated NOE assignment. Comparison with E2 structures present in the PDB and solved previously by X-ray crystallography or NMR shows in all cases the same compact fold. Differences are observed in the orientation of both the C-terminal α -helices and the N-terminal α -helix that is known to be involved in binding to the ubiquitin ligase (E3). In addition, the side chain of a conserved asparagine residue (Asn77), that is near the active site cysteine and that has been shown to be important for isopeptide bond formation, is solvent exposed. This positioning supports the proposed catalytic function of this amino acid (Wu et al., 2003).

Introduction

UbcH5B is a ubiquitin conjugating enzyme involved in the ubiquitination pathway, the main pathway for protein degradation in eukaryotes (Glickman and Ciechanover, 2002; Pickart, 2001; Weissman, 2001). In this pathway a ubiquitin is covalently attached to a substrate protein. The ubiquitinated protein is subsequently recognized and degraded by the 26S proteasome. UbcH5B is known to be essential for degradation of many regulatory and abnormal proteins (Seufert and Jentsch, 1990). Together with UbcH5A and UbcH5C, UbcH5B forms one of the most active class of E2 enzymes. It is associated with the degradation of a number of important human transcription factors, such as p53 (Scheffner et al., 1994), NF- κ B (Gonen et al., 1999), and c-fos (Stancovski et al., 1995). Recently it has been noted as well that ubiquitination of proteins also has other regulatory functions in for example signal transduction, transcription regulation, chromatin remodeling and DNA repair (Aguilar and Wendland, 2003). In all cases the attachment of ubiquitin moieties to the substrate is catalyzed by three enzymes: first, an E1, or ubiquitin activating enzyme, forms a thiol ester with the carboxyl terminal group of ubiquitin. Second, the ubiquitin is transferred to the ubiquitin conjugating enzyme (E2). Finally, an ubiquitin ligase (E3) transfers ubiquitin from E2 to the substrate protein. In human cells, one E1, about 30 E2s and at least 400 putative E3s have been identified. The selectivity and specificity of ubiquitination depends on the E2-E3 and E3-target complexes. Comparison of the different E2 and E3 structures and their complexes should shed light on the specificity encountered in such complex formation. Therefore structures of various E2s and E3s are needed. To date, 10 structures of E2 enzymes from different species have been solved in their free form, 9 by X-ray crystallography (Cook et al., 1992; Cook et al., 1993; Cook et al., 1997; Giraud et al., 1998; Hamilton et al., 2001; Jiang and Basavappa, 1999; Lin et al., 2002; Tong et al., 1997; VanDemark et al., 2001; Worthylake et al., 1998) and one by NMR (Miura et al., 2002). Furthermore, five structures of three different E2's in complex with various other proteins have been solved by X-ray crystallography (Bernier-Villamor et al., 2002;

Huang et al., 1999; Moraes et al., 2001; VanDemark et al., 2001; Zheng et al., 2000). All E2 structures possess the same compact fold corresponding to an N-terminal α -helix, followed by a four-stranded anti-parallel β -sheet and three C-terminal α -helices. Moreover the long stretch that contains the active site cysteine is very well-defined in all structures. The cysteine residue is solvent exposed, is located in a slight depression on the surface and is surrounded by loops. It has been previously postulated that the mechanism of transfer of ubiquitin from E1 to E2 and from E2 to the substrate resembles that of a thiol protease (Pickart, 2001). Based on the crystal structure of Ubc9 (Tong et al., 1997), some residues around the active site, among which one highly conserved asparagine residue (Asn77 in UbcH5B), have been proposed to catalyze the isopeptide bond formation. The importance of the asparagine residue was confirmed by mutagenesis experiments of three other E2s that revealed its necessity for efficient isopeptide bond formation in E2-catalyzed ubiquitin conjugation (Wu et al., 2003). In the different crystal structures of Ubcs, however, this asparagine is hydrogen-bonded to the loop connecting helix H2 and H3 and pointing away from the E2 cysteine and therefore can not participate in the enzymatic reaction. This implies that a repositioning of this asparagine residue must occur to catalyze the isopeptide bond formation (Tong et al., 1997; Wu et al., 2003).

Here we report the NMR structure of the ubiquitin conjugating enzyme UbcH5B using a combination of homology modeling, diffusion anisotropy restraints and automated NOE assignment. Dynamical properties of UbcH5B were assessed from ^{15}N relaxation measurements, which show limited motion for the major part of the protein backbone. The relaxation data (R_2/R_1) have been translated into diffusion anisotropy restraints and used in the automated NOE assignment and structure calculations (Bruschweiler et al., 1995; Tjandra et al., 1995; Tjandra et al., 1997). The final structure, which is well defined, possesses the canonical E2 fold, but differences in the position of the N- and C-terminal helices as compared to the core of the protein are observed. Since the N-terminal helix is involved in binding to ubiquitin ligase, the position of this helix in the different structures may be important for differentiating between various E3 ligases.

Results and Discussion

Assignment of UbcH5B and secondary structure prediction

The assignment of UbcH5B was previously reported (Farrow et al., 2000). However, the data were collected at different temperature, pH, and salt concentration than used here. Therefore, for the backbone assignment of UbcH5B in our conditions (300K, pH 7.0, 150 mM KCl), 2D (^{15}N - ^1H)-HSQC, 2D (^{13}C - ^1H)-HSQC, 3D HNCO, 3D HNCACB, and 3D CBCA(CO)NH were recorded. The side chain assignment was performed using 3D TOCSY- (^{15}N - ^1H)-HSQC, 3D H(C)CH-TOCSY and 3D (H)CCH-TOCSY spectra (for a review see Sattler et al., 1999). All residues except prolines were assigned. Finally, 85% of all observable protons could be assigned and a table with chemical shift is shown in Appendix 2 of this thesis. Dihedral angles of UbcH5B were predicted with the program

Talos (Cornilescu et al., 1999) based on the C α and C β chemical shifts. These predictions correlate well with the consensus secondary structure elements found in known E2 structures (data not shown).

Diffusion anisotropy and refinement of the homology model

The amide ^{15}N relaxation rates R_1 , R_2 and ^1H -NOE for 112 out of the 131 non proline residues of UbcH5B are shown in Figure 1. Residues that have overlapping peaks in the HSQC spectrum were discarded in the analysis, such as Glu122, which has only a very low intensity peak. UbcH5B has a rather rigid backbone, with an average heteronuclear NOE value of 0.78 ± 0.09 for all residues and 0.80 ± 0.08 for residues located in secondary structure elements. The average R_1 and R_2 values are $1.67 \pm 0.13 \text{ s}^{-1}$ and $10.5 \pm 1.1 \text{ s}^{-1}$, respectively, for the entire backbone and $1.68 \pm 0.12 \text{ s}^{-1}$ and $10.6 \pm 0.9 \text{ s}^{-1}$, respectively, for the secondary structure elements. The R_2/R_1 ratios, which are illustrative of the rotational diffusion characteristics of a protein, are presented in Figure 1D. The filled bars represent the 87 R_2/R_1 ratios that were selected for the determination of the rotational diffusion parameters of the protein. From the average R_2/R_1 ratio (6.3 ± 0.9) the isotropic diffusion tensor constant was estimated to be $1.9 \cdot 10^7 \pm 0.3 \cdot 10^7 \text{ s}^{-1}$, which gives an apparent correlation time of $8.6 \pm 1.2 \text{ ns}$. The histogram of R_2/R_1 ratios in Figure 2A, however, immediately reveals that UbcH5B does not tumble isotropically in solution. High R_2/R_1 ratios are found for the residues located in helix H2, caused by both their high R_2 rates ($> 11.6 \text{ s}^{-1}$) and low R_1 rates ($< 1.54 \text{ s}^{-1}$). The possibility that the high R_2 rates in this helix are caused by conformational exchange could be excluded, since no differences were observed between the ^{15}N R_2 rates determined from both CPMG and $T_{1\rho}$ experiments (data not shown). In addition relaxation dispersion profiles measured at both 500 and 700 MHz (Loria et al., 1999), that are very sensitive to conformational exchange, do not give any indication for such effects in helix H2 (data not shown). In a previous study of the interaction of UbcH5B with the RING domain of CNOT4 we generated a homology model of UbcH5B based on the structure of the highly homologous protein Ubc4 (Dominguez et al., 2004). In this model, helix H2 is approximately parallel to the long axis of the protein, which would explain the high R_2/R_1 ratios in this part of the protein. However, as shown in Figure 2B, the back-calculated R_2/R_1 values still deviate from the experimental values with a reduced χ^2 of 4.59 (for details, see Material and Methods). The deviation can be mainly attributed to the orientation of helix H2. The homology model was therefore refined in explicit water, using both TALOS-derived dihedral angles and diffusion anisotropy restraints. Three independent structure refinements were performed including the R_2/R_1 orientational restraints, starting from three different estimates of the anisotropy A and rhombicity η of the diffusion tensor: i) 1.61 and 0.33, ii) 1.68 and 0.27 and iii) 1.64 and 0.31 which were obtained as described in the Material and Methods section. The average reduced χ^2 value decreased from 4.59 to 2.20 ± 0.13 , 2.21 ± 0.19 and 2.44 ± 0.70 for the three refined sets each consisting of twenty models. Using these sixty refined models a grid search was performed to optimize the diffusion tensor components, resulting in an anisotropy A of 1.68 and a rhombicity η of 0.30 with a reduced χ^2 of 2.34 ± 0.38 . These were used as starting values for the diffusion tensor in

all subsequent structure calculations. The average pairwise backbone RMSD between the sixty refined models is $0.7 \pm 0.1 \text{ \AA}$ and the backbone RMSD from the original homology model is $0.84 \pm 0.02 \text{ \AA}$.

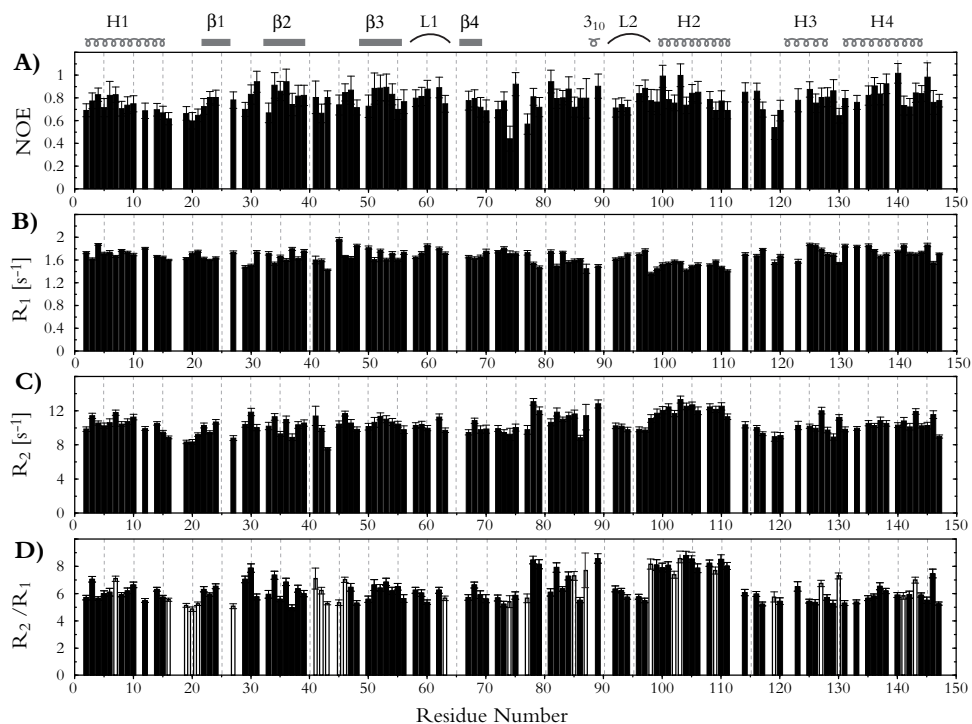


Figure 1: ^{15}N relaxation parameters of UbcH5B at 300K and 500MHz. **A)** Heteronuclear NOE values, **B)** R_1 relaxation rates, **C)** R_2 relaxation rates from CPMG experiments and **D)** R_2/R_1 ratios. The filled bars in **D)** represent the 87 selected ratios. The secondary structure elements with the two loops L1 and L2 involved in the interaction with the CNOT4 E3 ligase (Dominguez et al., 2004) are indicated on top.

Model-based automated NOE assignment and structure calculations of UbcH5B

In order to facilitate and speed up structure determination by NMR, approaches have been developed that combine structure calculations and automated assignment of NOE spectra in an iterative manner (Guntert, 2003; Herrmann et al., 2002; Linge et al., 2001). For large proteins however, the ambiguity in the assignment increases rapidly due to peak overlap in the spectra and makes automated assignment in that case difficult, causing poor convergence of the calculated structures to an unique fold. It has been suggested that the use of starting models can assist and speed up automated NOE assignment procedures (Duggan et al., 2001; Guntert et al., 1993; Hare and Wagner, 1999).

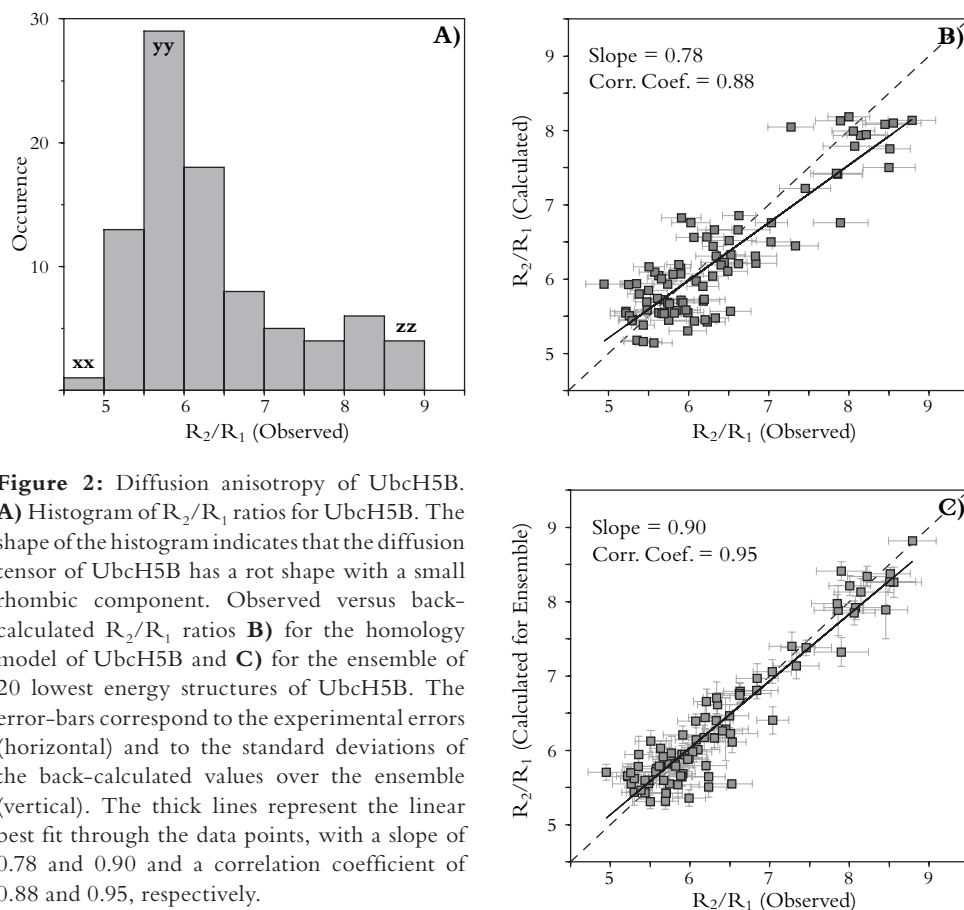


Figure 2: Diffusion anisotropy of UbcH5B. **A)** Histogram of R_2/R_1 ratios for UbcH5B. The shape of the histogram indicates that the diffusion tensor of UbcH5B has a rot shape with a small rhombic component. Observed versus back-calculated R_2/R_1 ratios **B)** for the homology model of UbcH5B and **C)** for the ensemble of 20 lowest energy structures of UbcH5B. The error-bars correspond to the experimental errors (horizontal) and to the standard deviations of the back-calculated values over the ensemble (vertical). The thick lines represent the linear best fit through the data points, with a slope of 0.78 and 0.90 and a correlation coefficient of 0.88 and 0.95, respectively.

The automated NOE assignment and structure calculations of UbcH5B were performed with a version of ARIA1.2 (Linge et al., 2001) that we modified to include diffusion anisotropy restraints (see Material and Methods). This modified version of ARIA has been made available via the ARIA homepage (<http://www.pasteur.fr/recherche/unites/Binfs/aria>). A total of 5931 NOEs were obtained from a 2D NOE and a 3D (^{15}N - ^1H)-NOESY-HSQC spectra recorded on a 900 MHz spectrometer. From the 3D (^{15}N - ^1H)-NOESY-HSQC and the 2D NOE spectra 1002 and 4929 (on both sides of the diagonal) NOE peaks were extracted, respectively. 404 intra and 284 sequential NOEs were manually assigned between the amide and the side chains protons for the 2D NOE spectrum and 350 intra and 248 sequential for the 3D spectrum. The chemical shift of the corresponding nuclei were adjusted based on these assignments, which allowed the use of tighter chemical shift value tolerances in the automated NOE assignment protocol: 0.02 ppm for the ^1H chemical shift values and 0.2 ppm in the ^{15}N dimension. In addition, 76 dihedral angle restraints derived

from TALOS (Cornilescu et al., 1999) and 87 diffusion anisotropy restraints were used in the structure calculation.

Three successive ARIA runs (run1-3) were performed, each consisting of nine iterations and a final refinement in explicit water for the last run. In run1, the 60 refined models were used to perform the automated assignment of the NOEs. No structures were calculated at this level yet. The distance violation tolerance was decreased in the successive iterations from 5 to 0.3 Å with a constant value of 0.90 for the ambiguity cut-off. The number of assigned unambiguous NOEs was the highest for a tolerance of 0.6 Å (1680 unambiguous and 1630 ambiguous NOEs after merging of the two spectra). Based on these assignments two subsequent runs of structure calculations were performed including TALOS derived dihedral angles and R_2/R_1 derived diffusion anisotropy restraints. Run2 started from an extended structure, using the set of 1680 unambiguous and 1630 ambiguous NOE restraints from run1. Run2 consisted of 9 ARIA iterations, in which the structures gradually converged and in which these NOE restraints could be reassigned, recalibrated and rejected if violated in more than 50% of the 20 best structures of the previous iteration. For the last iteration the anisotropy A and rhombicity η of the diffusion tensor were again optimized using a grid search resulting in values of $A = 1.72$ and $\eta = 0.27$. This run resulted in 1902 unambiguous and 1145 ambiguous NOE restraints and in an already well-defined ensemble of structures (average backbone RMSD from the mean structure of 0.86 ± 0.22 Å for the secondary structure elements). In the final run, run3, the 1902 unambiguous and 1145 ambiguous NOE restraints of run2 were fixed. In addition the NOE peaks that were not previously converted into either unambiguous or ambiguous restraints were added into the calculation. This consisted of a total number of 2318 NOE peaks. The initial automated NOE assignment was based on the 20 final structures of run2. The optimized values of A and η from run2 were used for the diffusion anisotropy restraints. After each iteration the diffusion tensor was optimized, resulting in a final value of 1.76 and 0.22 for A and η , respectively. This finally resulted in a total of 2049 unique unambiguous and 1567 unique ambiguous NOE restraints (Table 1) derived from 4092 and 859 assigned cross-peaks in the NOESY and 3D-NOESY-HSQC spectra, respectively. Figure 3A shows the distribution of the unambiguous NOEs over the polypeptide sequence. In general a large number of NOE is found in the secondary structure elements.

After water refinement, the average backbone RMSD from the mean for the secondary structure elements of the 20 lowest energy structures is 0.56 ± 0.1 Å (all heavy atoms: 1.04 ± 0.12 Å) (Table 1). Figure 3B clearly shows that the RMSD values in the loops between the secondary structure elements are significantly higher. This clearly correlates with the lower number of NOEs found in those regions. The structural statistics are presented in Table 1. The average backbone RMSD between the ensemble of structures and the original homology model is 2.32 ± 0.09 Å whereas for the secondary structures, the backbone RMSD is 2.06 ± 0.09 Å. The average reduced χ^2 for the diffusion anisotropy restraints is 2.93 ± 0.40 for the ensemble and 2.44 for the representative structure. Experimental against back-calculated R_2/R_1 ratios are plotted in Figure 2C. The correlation coefficient of 0.95 indicates a good agreement between the calculated and experimental values. The

large number of NOEs result in a reduced χ^2 which is slightly higher than for the refined homology models (see above). The latter would however violate more than 800 NOE restraints.

Table 1: Structural statistics of the 20 best UbcH5B structures.

Number of experimental restraints:	
Intra-residue unambiguous NOEs	847
Sequential unambiguous NOEs	442
Medium-range unambiguous NOEs	301
long-range unambiguous NOEs	459
Total unambiguous NOEs	2049
Total ambiguous NOEs	1567
Dihedral angles ^a	152 (76 ϕ +76 ψ)
Diffusion anisotropy ^b	87
R.m.s.d. (Å) from the mean	
All backbone atoms	0.68 ± 0.12
All heavy atoms	1.13 ± 0.13
Secondary structure backbone atoms ^c	0.56 ± 0.10
Secondary structure heavy atoms ^c	1.04 ± 0.12
Non-bonded energy values ^d after water-refinement (kcal mol ⁻¹)	
E vdW	-1305 ± 133
E electrostatic	-5043 ± 90
R.m.s.d from idealized covalent geometry	
Bonds (Å)	0.0063 ± 0.0001
Angles (°)	0.80 ± 0.01
Impropers (°)	2.23 ± 0.07
R.m.s.d from experimental data	
Distance (Å)	0.13 ± 0.01
Dihedral (°)	3.98 ± 0.85
Diffusion anisotropy	0.37 ± 0.02
Restraint violations in more than 50% of the structures ^e	
Distance (> 0.3 Å)	2
Dihedral (> 5°)	4
Diffusion anisotropy (> 0.7)	2
Ramachandran analysis	
Residues in most favored region (%)	72.8 ± 2.4
Residues in additional allowed regions (%)	21.1 ± 2.2
Residues in generously allowed regions (%)	4.6 ± 2.1
Residues in disallowed regions (%)	1.4 ± 0.1

a) The dihedral angles are derived from Talos (Cornilescu et al., 1999) prediction based on C α and C β chemical shifts.

b) The diffusion anisotropy restraints are derived from the R₂/R₁ ratios.

c) Secondary structure elements comprise residues 2-15, 21-24, 32-38, 49-55, 66-70, 99-111, 122-128, and 131-144.

d) The non-bonded energies were calculated with the OPLS parameters using a 8.5 Å cutoff.

e) No NOE distance restraint was violated by more than 0.65 Å, no dihedral angles by more than 6.8° and no diffusion anisotropy by more than 0.98.

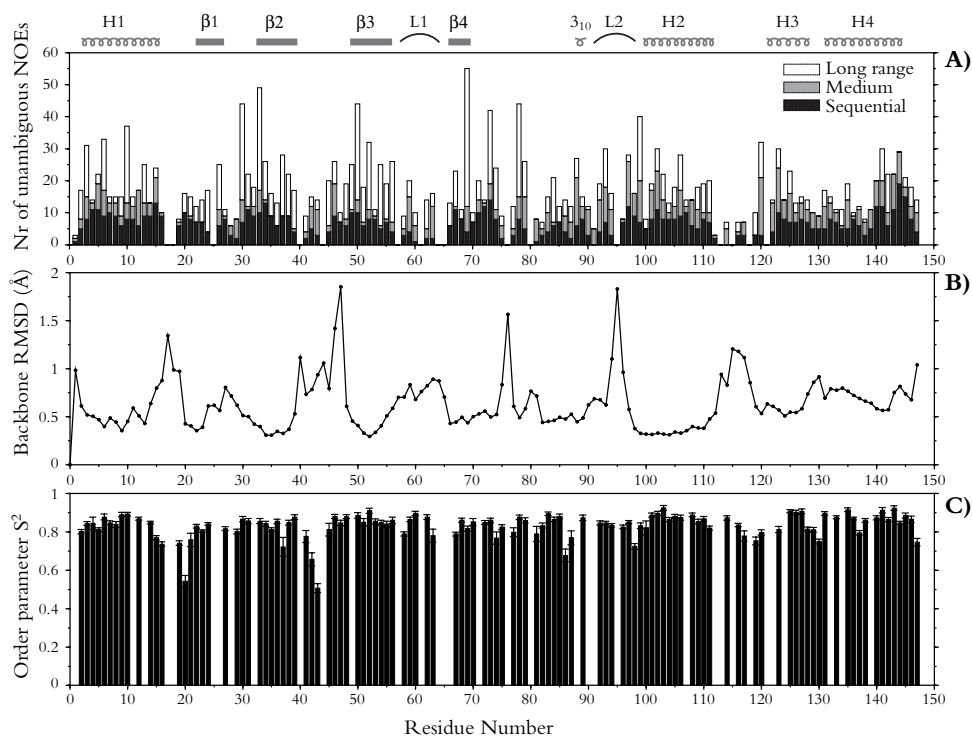


Figure 3: Structural and dynamical data of the UbcH5B NMR structure. **A)** Distribution of the number of unambiguous NOE restraints over the protein sequence. The number indicated are the final numbers used in the structure calculation. **B)** Average backbone RMSD for each residue from the mean. The values shown are the average RMSD calculated from the ensemble of 20 structures. **C)** Generalized order parameter S^2 obtained from the *model-free* analysis of the ^{15}N relaxation data. Secondary structure elements and loops L1 and L2 are indicated at the top of the figure.

Solution structure of UbcH5B

The structure of UbcH5B is composed of a four-stranded anti-parallel β -sheet and four α -helices. The N-terminal α -helix (residues 2-15) is followed by the four-stranded β -sheet (residues 21-24, 32-38, 49-55, 66-70). A long extended stretch (residues 71-86), a short 3_{10} helix (residues 87-89) and a loop (residues 90-98) connect the last β -strand to the second α -helix (residues 99-111). The second α -helix (H2) is parallel to the long axis of the proteins in agreement with the diffusion anisotropy data. The C-terminal part of the protein is composed of two α -helices (residues 122-128 and 131-144) (Figure 4). The core of the protein, which consists of the β -sheet, the long extended stretch that contains the active site cysteine and the α -helix H2 is very well defined with an average backbone RMSD from the mean of 0.32 ± 0.09 Å (all heavy atoms, 0.87 ± 0.08 Å). The N-terminal α -helix is almost as well defined as the core of the protein. The two C-terminal α -helices are slightly more disordered probably due to a low number of NOEs in this region.

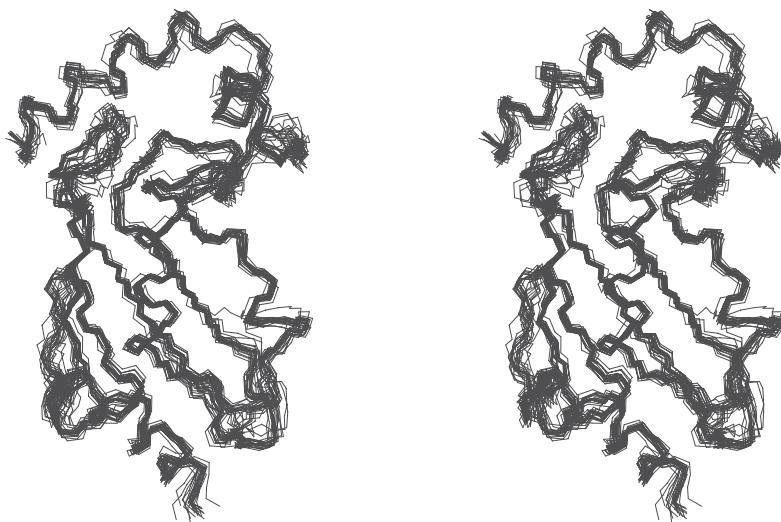


Figure 4: NMR structure of UbcH5B. Stereoview of the 20 best structures. This figure was generated using the program MOLMOL (Koradi et al., 1996).

Backbone dynamics of UbcH5B

The average anisotropic rotational diffusion tensor for the ensemble of 20 UbcH5B structures was analyzed with the program TENSOR2 (Dosset et al., 2000). This resulted in an overall rotational correlation time τ_c of 8.58 ± 0.02 ns, an anisotropy A of 1.75 ± 0.02 and a rhombicity η of 0.17 ± 0.04 . The overall rotational correlation time corresponds nicely with the correlation time expected for a monomeric protein of this size at 300K and with the value of 8.5 ns calculated by the program HYDRONMR (Bernado et al., 2002; de la Torre et al., 2000). Using the determined anisotropic diffusion tensor the ^{15}N relaxation rates (Figure 1) were analyzed in TENSOR2 using the *model-free* (Lipari and Szabo, 1982) approach based on the representative structure. The resulting order parameters S^2 are shown in Figure 3C and shaded on the structure in Figure 5. The average order parameter is 0.83 ± 0.07 . The most flexible parts in UbcH5B are the loop connecting the N-terminal helix H1 with the β -sheet and the loop connecting the strands $\beta 2$ and $\beta 3$. Some residues with lower order parameters are also present in the long extended stretch that contains the active site cysteine. In the loop L2, which is known to be involved in binding to E3 ligases (Zheng et al., 2000), residue Thr98 has a lower order parameter. Furthermore, flexible residues are found in the loop between helix H2 and H3. This loop forms a 'lid' on top of the active site Cys85 with in particular the side chain of Leu119 in close proximity to the cysteine side chain. The flexibility in this loop could therefore affect both ubiquitin binding and ubiquitination of the substrate. In the C-terminal part of the protein both Asp130, located between the helices H3 and H4, and the C-terminal residue Met147 have a lower order parameter.

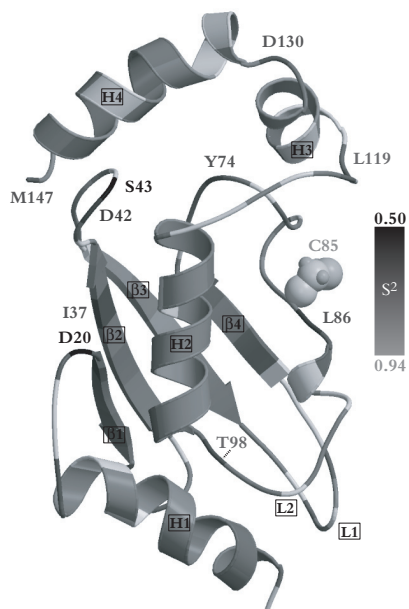


Figure 5: Dynamics in UbcH5B. The ribbon representation of UbcH5B is shaded according to the S^2 values derived from the ^{15}N relaxation analysis. The grey-scaling is from grey for high S^2 values to black for low S^2 values. Residues with low order parameters are labeled. Residues for which no S^2 was determined are white. The active site cysteine 85 is displayed in a space-filling representation. The α -helices (H1 to H4), β -strands ($\beta 1$ to $\beta 4$) and the two loops (L1 and L2) involved in E2 binding are labeled. This figure was generated with the programs Molscrip (Kraulis, 1991) and Raster3D (Merritt and Murphy, 1994).

Comparison of UbcH5B with other Ubc enzymes

For comparing the structure and dynamics of UbcH5B with other Ubc enzymes, we focus on Ubc4 (Cook et al., 1993), which is highly homologous to UbcH5B and on Ubc9 (Giraud et al., 1998; Tong et al., 1997) of which the backbone dynamic has been studied by NMR before (Liu et al., 1999). The core domain of UbcH5B is very similar to the X-ray structure of Ubc4 (backbone RMSD of 0.99 Å) (Figure 6). This close structural resemblance of the core domain is a common feature of all Ubc structures, which all have a RMSD within 1.3 Å for this part. It is interesting to note that the 3_{10} helix (residues 87–89) is conserved among all E2s and that the long extended stretch that contains the active site cysteine (residues 82 to 89 in UbcH5B) has a very similar conformation in all Ubc structures, in particular at the active site cysteine. Clear differences in dynamics between UbcH5B and Ubc9 within this extended stretch can be observed, however, on the basis of the ^{15}N relaxation experiments. In Ubc9 Leu81, Val86, Ser89 and Leu97 have low order parameters and for both Val86 and Leu97 a substantial contribution of conformational exchange R_{ex} to the ^{15}N R_2 relaxation rate was detected. The corresponding residues in UbcH5B (Ile73, Ile78, Asn81 and Leu89) do not have low order parameters. The residues that do have lower order parameters in UbcH5B in this region are Tyr74 and Leu86. Although UbcH5B and Ubc9 are homologous proteins, their biological function is quite distinct. Whereas UbcH5B is primarily involved in ubiquitination of proteins, Ubc9 plays an important role in sumoylation (Muller et al., 2001). Since the differences in flexibility are mainly on the face of the protein where the active site Cys85 is located, this could reflect this different biological function.

Mutagenesis data previously showed that a highly conserved asparagine is important for

the efficient catalytic activity of E2s in transferring ubiquitin to the substrate (Wu et al., 2003). In all existing X-ray structures the side chain of this asparagine is hydrogen-bonded to the backbone of the loop connecting the helices H2 and H3, and is thus pointing away from the active site cysteine (Tong et al., 1997; Wu et al., 2003). A structural rearrangement should therefore occur upon ubiquitination of the substrate. In the solution structure of UbcH5B, however, the corresponding Asn77 side chain, which is not hydrogen-bonded to the backbone of the loop connecting the helices H2 and H3, is already solvent exposed and in close proximity (within 4 Å) to the active site cysteine. Its orientation in solution is thus suitable for isopeptide bond formation (Figure 7). A suitable similar orientation of a corresponding asparagine (Asn80) was observed in the only other E2 (UbcH2B) for which the structure has been solved in solution as well (Miura et al., 2002). The absence of an hydrogen bond between the Asn77 side chain and the backbone of the loop between H2 and H3 is supported by the presence of flexibility in the latter loop. Indeed Asp117, Leu119 and Val120 have lower order parameters than average (the two other residues in this loop are prolines). In addition, Glu122 at the start of helix H3, for which no reliable relaxation parameters could be obtained, shows a low intensity peak in the HSQC spectrum, which could be an indication for the presence of conformational exchange. The same loop was also shown to be flexible in the NMR relaxation studies of Ubc9 (Liu et al., 1999).

Figure 6: Overlay of the UbcH5B and Ubc4 structures (PDB: 1QCQ). The figure shows the differential position of helix 1 in the two structures. UbcH5B is displayed black and Ubc4 white. This figure was generated with the programs Molscript (Kraulis, 1991) and Raster3D (Merritt and Murphy, 1994).



Other differences between the various Ubc structures are observed in the position of the first α -helix and the last two α -helices. Here we find a relatively high backbone RMSD of 2.29 Å between UbcH5B and Ubc4. This seems to be a general feature of the Ubc family since most differences are observed for these three helices. The first N-terminal helix is interesting since it is part of the interacting site with ubiquitin ligases (both HECT

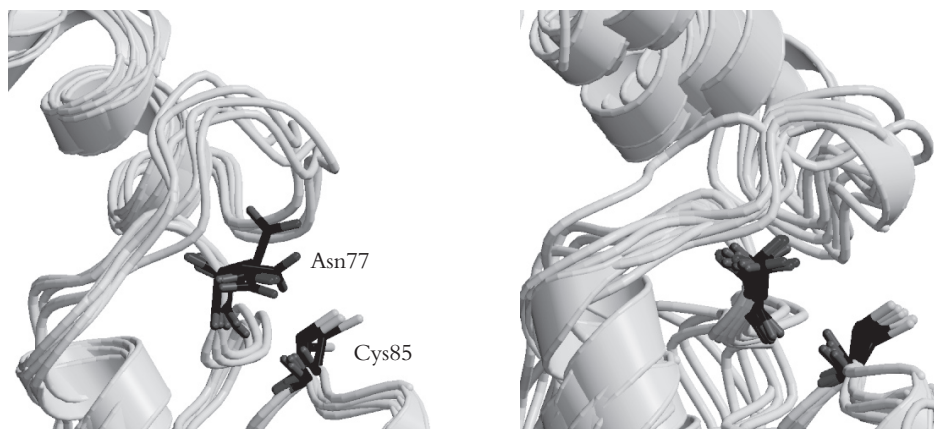


Figure 7: Comparison of the position of the asparagine residue important for oxyanion stabilization in E2-catalyzed ubiquitin conjugation. Right: NMR ensemble of UbcH5B (5 structures). Left: Overlay of 8 E2 X-ray structures (PDB-codes: 1QCQ, 1A3S, 1AYZ, 1FZY, 1U9A, 2AAK, 2UCZ, 2EZC). This figure was generated with the programs Molscript (Kraulis, 1991) and Raster3D (Merritt and Murphy, 1994).

and RING finger domains) (Dominguez et al., 2004; Huang et al., 1999; Zheng et al., 2000). In the UbcH5B structure many NOEs are present between the residues of this helix and residues of loop L1 (residues 56–65) and L2 (residues 89–98), which are also part of the interacting site. This helix makes thus multiple contacts with these two loops. In the structure of UbcH5B, the first helix is in closer proximity to loops L1 and L2 than in the Ubc4 structure. The differential position of this helix with respect to the loops L1 and L2 may contribute to the recognition of various E3 ligases. The loop connecting the first helix (H1) and the β -sheet is flexible in both UbcH5B and Ubc9. This may allow for structural rearrangements of helix H1 relative to the rather rigid L1 and L2 loops to accommodate binding to different E3s. Using NMR chemical shift perturbation experiments, we already reported that residues in helix H1 (Ala2, Leu3, Arg5, Ile6, Glu9, Leu10 and Asp12) and loop L2 (Thr98, Ile99) are affected by the binding to CNOT4 (Dominguez et al., 2004). These residues are, however, not solvent exposed but responsible for the H1-L2 interaction. The chemical shift changes of these residues could therefore be due to the displacement of the first helix. It is therefore likely that, during binding, E2 enzymes undergo structural rearrangements involving the first helix to accommodate the different E3 ligases.

Conclusions

We have described here the solution structure and dynamics of the human UbcH5B, an E2 enzyme involved in the ubiquitination pathway. Comparison of this structure with

other E2 structures shows striking differences, mainly at the N- and C-terminal parts of the proteins. These observed differences may be important for the recognition of different partners. Moreover, flexibility was observed in specific loops of the protein, especially close to the active site cysteine and in the N-terminal domain. This flexibility of E2s may be important to allow for small structural rearrangements when binding to different partners. In the surrounding of the active site cysteine, a conserved asparagine residue has been shown to be important in the catalysis of the isopeptide bond formation during the transfer of ubiquitin to the substrate. In our solution structure, this residue is solvent exposed and in close proximity to the active site cysteine, in a position thus suitable for catalysis.

Material and Methods

Recombinant protein expression and purification

The construction of the plasmid for the expression of the human UbcH5B and the overexpression of ^{15}N and $^{13}\text{C}/^{15}\text{N}$ isotopically labeled UbcH5B have been described previously (Albert et al., 2002; Dominguez et al., 2004). The UbcH5B samples were finally concentrated to a final concentration of approximately 0.5 mM in an NMR buffer (150 mM KCl, 20 mM KPi pH 7.0, 10 μM ZnCl_2).

Homology modeling

The homology model of UbcH5B is based on the structure of yeast Ubc4 (pdb: 1QCQ) (Cook et al., 1993) and is generated using Modeller4 (Sali and Blundell, 1993). For details see our previous study on the interaction between UbcH5B and the RING domain of CNOT4 (Dominguez et al., 2004). This model was used here as a starting point for automated NOE assignment and structure calculations of UbcH5B in ARIA1.2 (Nilges et al., 1997).

NMR measurements

For the backbone assignment of UbcH5B, 2D (^{15}N - ^1H)-HSQC, 2D (^{13}C - ^1H)-HSQC, 3D HNCQ, 3D HNCACB, and 3D CBCA(CO)NH were recorded on a Bruker AVANCE 700 MHz spectrometer. The side chain assignment was performed using 3D TOCSY-(^{15}N - ^1H)-HSQC, 3D H(C)CH-TOCSY and 3D (H)CCH-TOCSY spectra also recorded on a Bruker AVANCE 700 MHz spectrometer (for a review see Sattler et al., 1999).

The NOESY spectra were recorded on a Bruker AVANCE 900 MHz spectrometer. The NOE information was extracted from a 3D NOESY-(^{15}N - ^1H)-HSQC and a NOESY with mixing times of 100 ms. The NOESY was recorded with 768 complex points in the direct dimension (spectral width of 15151 Hz) and 384 complex points in the indirect dimension (spectral width of 12820 Hz). The 3D NOESY-(^{15}N - ^1H)-HSQC was recorded with 768 complex points in the direct dimension (spectral width of 15151 Hz), 160 complex points in the ^1H indirect dimension (spectral width of 15151 Hz), and 40 complex points in the ^{15}N dimension (spectral width of 3333 Hz). For Fourier transformation all dimensions were zero-filled twice.

All relaxation experiments were performed at 300K on a Bruker AVANCE 500 MHz spectrometer (^1H frequency of 500.28 MHz) equipped with a QXI probe with z-gradients using a ^{15}N -labeled UbcH5B sample with a concentration of approximately 0.5 mM. ^{15}N T_1 and heteronuclear $\{^1\text{H}\}$ -NOE values were determined using the experiments described by Farrow et al., (1994). T_1 times were extracted from eight spectra with different values for the relaxation delay: 100 (2x), 200, 300, 400 (2x), 500, 600, 800 and 1000 ms, giving 180° pulses on proton every 5 ms to suppress cross-correlated relaxation. The heteronuclear NOE was recorded in an interleaved fashion, recording alternatively one increment for the reference and one for the NOE spectrum. In the NOE experiment the protons were saturated using 120° pulses (20.7 kHz). ^{15}N T_2 relaxation times were extracted from both CPMG (Carr and Purcell, 1954; Meiboom and Gill, 1958) and $T_{1\rho}$ (Peng et al., 1991) experiments. CPMG experiments were recorded using nine different values for the relaxation delay: 0 (2x), 16.1 (2x), 32.2, 48.2, 64.3 (2x), 80.4, 96.5, 128.6 and 160.8 ms. During the relaxation delay ^{15}N 180° pulses with a field strength of 7.8 kHz were applied every 0.95 ms ($\nu_{\text{CPMG}} = 1$ kHz) and ^1H 180° pulses were applied every 7.7 ms, to suppress cross-correlated relaxation pathways (Kay et al., 1992). The $T_{1\rho}$ experiments were recorded with varying lengths of the spin-lock pulse: 2, 4, 6, 8, 10 (2x), 20, 30, 50 (2x), 70, 100, 150 ms. An adiabatic spin-lock pulse, as described by Mulder et al., (1998), was used to align the magnetization of the individual amides along their effective field. The pulse was applied on-resonance with a field-strength of 2.5 kHz. The number of ^1H 180° pulses during the relaxation period was adapted to the used relaxation delay (Korzhnev et al., 2002). No ^1H 180° pulses were applied up to 30 ms, one ^1H 180° pulse was applied in the middle of the 50 and 70 ms delays, two pulses in the case of 100 ms and three for the 150 ms delay.

All spectra were processed with NMRPipe (Delaglio et al., 1995) and analyzed using NMRView5.0.4 (Johnson and Blevins, 1994). Relaxation parameters were extracted and analyzed with the program Curvefit (Palmer et al., 1991): (<http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software/curvefit.html>), using a 2-parameter fitting and a Monte Carlo simulation to estimate the errors.

Diffusion anisotropy

In the absence of large amplitude internal motions and conformational exchange, the ratio of the ^{15}N transverse and longitudinal relaxation rate (R_2/R_1) is dependent on the angles θ and ϕ between the amide bond vector and the diffusion tensor of the protein. Given a good estimate of the diffusion tensor components (D_{xx} , D_{yy} and D_{zz}), the R_2/R_1 ratios can thus be used to refine the orientation of the amide bond vectors (Tjandra et al., 1997). These ratios are included in CNS (Brunger et al., 1998) as diffusion anisotropy restraints (DANI) as described by Tjandra et al., (1997). The diffusion tensor components are defined by the overall rotational correlation time (τ_c), the anisotropy (A) and the rhombicity (η):

$$\tau_c = \frac{1}{2 \cdot (D_{xx} + D_{yy} + D_{zz})} \quad (1)$$

$$A = \frac{2D_{zz}}{(D_{yy} + D_{xx})} \quad (2)$$

$$\eta = \frac{1.5 \cdot (D_{yy} - D_{xx})}{D_{zz} - 0.5 \cdot (D_{yy} + D_{xx})} \quad (3)$$

In CNS (Brunger et al., 1998) the geometric content of the R_2/R_1 ratios is incorporated in the simulated annealing protocol for the structure calculation by minimizing the harmonic potential energy term E_{dani}

$$E_{\text{dani}} = k_{\text{dani}} \sum \left\{ \left[(R_2/R_1)_{\text{calc}} - (R_2/R_1)_{\text{obs}} \right]^2 \right\} \quad (4)$$

where k_{dani} is the force constant for the diffusion anisotropy restraints and $(R_2/R_1)_{\text{calc}}$ and $(R_2/R_1)_{\text{obs}}$ are the back-calculated and observed R_2/R_1 ratios, respectively.

The selection of R_2/R_1 ratios to determine the diffusion tensor components and to define diffusion anisotropy restraints is as follows. All residues are selected that have both an heteronuclear NOE value higher than 0.65 and a R_2 rate smaller than the average rate plus one standard deviation. In addition, residues with a high R_2 rate and a corresponding R_1 rate lower than the average rate minus one standard deviation are also selected. This selection procedure keeps the residues with high R_2 values caused by the anisotropy of the system (Pawley et al., 2001). In this way, 87 R_2/R_1 ratios were selected. The diffusion tensor components were determined in three different ways:

- i) From the average of the highest, the lowest and the most occurring R_2/R_1 ratios in the distribution (Figure 2A) the tensor components can be estimated as has been described by Clore *et al.*, (1998). This gave a value of 1.61 for the anisotropy A and 0.33 for the rhombicity η .
- ii) Since an average of the extreme values is used in the first approach, the anisotropy is rather under- than overestimated. Therefore a more extreme estimate, using the averages of the two minimum and maximum values minus and plus the standard deviation, respectively, gives a value of 1.68 for A with a corresponding η of 0.27.
- iii) When a proper structural model is available, TENSOR2 (Dosset et al., 2000) can be used to determine the rotational diffusion tensor, based on the selected set of relaxation rates. Using the homology model of UbcH5B the computed values are 1.64 for A and 0.31 for η , respectively. These three sets of values for the diffusion tensor were used to refine the homology model of UbcH5B (see below).

The rotational diffusion parameters of a molecule can also be estimated by hydrodynamic modeling. This has been implemented in the program HYDRONMR (de la Torre et al., 2000), which uses a bead shell method to perform the hydrodynamic calculations. An important parameter in these calculations is the radius a of the spherical elements that are used to replace each nonhydrogen atom to build the initial shell model. For most proteins this atomic element radius a has a value between 2.8 Å and 3.8 Å, with a distribution centered at 3.3 Å (Bernado et al., 2002). To calculate the hydrodynamic properties of

UbcH5B we used a value of 3.2 Å for a and a solvent viscosity of 0.8 cP at a temperature of 300 K.

The agreement between back-calculated and experimental R_2/R_1 ratios can be expressed by the reduced χ^2 :

$$\chi^2 = \frac{1}{N} \sum \left\{ \frac{[(R_2/R_1)_{obs} - (R_2/R_1)_{calc}]^2}{\sigma_{R_2/R_1}^2} \right\} \quad (5)$$

where σ is the error in the ratios and N the number of observables.

Structure calculation

The automated assignment and structure calculations of UbcH5B were performed with ARIA1.2 (Linge et al., 2001) using CNS (Brunger et al., 1998). The topallhdg5.3.pro (Linge et al., 2003) topology and parameter set was used based on the PROLSQ parameters (Engl and Huber, 1991). The ARIA1.2 scripts were modified to allow the use of diffusion anisotropy restraints as defined in CNS (Tjandra et al., 1997). A grid search procedure to optimize the initial values for the anisotropic (A) and rhombic (η) components of the diffusion tensor, based on the calculated structures, was introduced. Herein the sum of the restraint energy term E_{dani} over the ensemble of structures is minimized by a grid search of ± 0.1 and ± 0.05 around the starting values of the diffusion tensor components A and η , in steps of 0.02 and 0.01, respectively.

First, the homology model of UbcH5B was refined using both TALOS dihedral and diffusion anisotropy restraints in explicit water. This refinement was performed starting from three different sets of estimated values of the rotational diffusion tensor, resulting in three ensembles of each 20 structures. The force constants for the diffusion anisotropy and TALOS dihedral angle restraints were set to 10 kcal mol⁻¹ and 200 kcal mol⁻¹ rad⁻², respectively. Based on the resulting 60 models the diffusion tensor components were optimized using the grid search described above and the resulting values were used in the subsequent structure calculation runs. The automated NOE assignment was performed in three steps as described below.

i) First the 60 models were used to create an initial set of NOE assignments, without calculating any structure. Because of the sensitivity of the 900 MHz spectrometer and the size of the protein, the upper bound limit for the NOE calibration was set to 7 Å and spin diffusion correction (Linge et al., 2004) was used in all runs.

ii) In the second run, consisting of 9 ARIA iterations, the assigned peaks from the first run were used to calculate an initial ensemble of structures, including both diffusion anisotropy and TALOS dihedral restraints. After each iteration the NOEs were reassigned and recalibrated based on the 20 lowest energy structures and rejected if violated in more than 50% of the structures. The ambiguity cut-off was reduced from 1.01 in the first iteration to a final value of 0.90 in the last iteration. The violation tolerance was set to 1000 Å for the first two iterations and then progressively reduced to 0.1 Å in the last iteration. In the last two iterations, the tensor components were optimized using a grid search as described above. For the structure calculations a simulated annealing (SA) protocol consisting of four stages was

used starting from an extended conformation using both torsion angle dynamics (TAD) and Cartesian dynamics. i) The high temperature TAD stage consisted of 10000 steps at 10000 K. This was followed by ii) a 8000 steps TAD cooling stage with a final temperature of 2000 K, iii) a 5000 steps first Cartesian cooling stage to 1000 K, and iv) a 10000 steps second Cartesian cooling stage to 50 K. During the SA protocol the force constants for the TAD stages, first Cartesian cooling stage and second Cartesian cooling stage were set to 10, 10 and 50 kcal mol⁻¹ Å⁻² for the NOE restraints, to 50, 100 and 200 kcal mol⁻¹ rad⁻² for the dihedral restraints and to 1, 1 and 10 kcal mol⁻¹ for the diffusion anisotropy restraints, respectively. The number of calculated structures in the iterations was 100 (50) for iterations 1 and 2, 20 (20) for iterations 3 to 7, 50 (20) for iteration 8 and 100 (50) for the final iteration, with the structures that were kept in the subsequent iteration indicated between brackets.

iii) In the third and last run the final ensemble of structures was calculated, starting in the first iteration from the ensemble of 20 lowest energy structures of run2. The NOE assignments of run2 were used and kept fixed. In addition the peak lists containing the remaining unassigned NOEs were introduced, to allow the assignment of more NOEs. The ambiguity cut-off was 0.90 in all iterations and the violation tolerance was reduced from 1.0 to 0.1 Å during the nine iterations, with only an increased value of 1.0 for iteration 6. The SA protocol described above was used for the structure calculations. The number of calculated structures was 50 (25) for iteration 2 to 8 and 100 (50) for the last iteration, with the structures that were kept in the subsequent iteration between brackets. After each iteration an optimization of the diffusion tensor components was performed, based on the ensemble of 20 lowest energy structures. The 50 final lowest energy structures were refined in explicit water using the OPLS parameters (Jorgensen and Tirado-rives, 1988).

Acknowledgements

This investigation was supported by the Netherland Foundation for Chemical Research (NWO-CW) and the Netherlands Organization for Scientific Research (NWO). This work was also supported by NWO “Jonge Chemici” and “Pionier” grants to A.M.J.J. Bonvin, and H. Th. M. Timmers, respectively. C. Dominguez was supported financially by the Center for Biomedical Genetics. The authors are grateful to Aart Nederveen for helpful discussion.

References

- Aguilar, R. C., and Wendland, B. (2003). Ubiquitin: not just for proteasomes anymore. *Curr Op Cell Biol* 15, 184-190.
- Albert, T. K., Hanzawa, H., Legtenberg, Y. I. A., de Ruwe, M. J., van den Heuvel, F. A. J., Collart, M. A., Boelens, R., and Timmers, H. T. M. (2002). Identification of a ubiquitin-protein ligase subunit within the CCR4-NOT transcription repressor complex. *EMBO J* 21, 355-364.

- Bernado, P., de la Torre, J. G., and Pons, M. (2002). Interpretation of N-15 NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. *J Biomol NMR* *23*, 139-150.
- Bernier-Villamor, V., Sampson, D. A., Matunis, M. J., and Lima, C. D. (2002). Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* *108*, 345-356.
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallog D* *54 (Pt 5)*, 905-921.
- Bruschweiler, R., Liao, X. B., and Wright, P. E. (1995). Long-Range Motional Restrictions in a Multidomain Zinc-Finger Protein from Anisotropic Tumbling. *Science* *268*, 886-889.
- Carr, H. Y., and Purcell, E. M. (1954). Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys Rev* *94*, 630-638.
- Clore, G. M., Gronenborn, A. M., Szabo, A., and Tjandra, N. (1998). Determining the magnitude of the fully asymmetric diffusion tensor from heteronuclear relaxation data in the absence of structural information. *J Am Chem Soc* *120*, 4889-4890.
- Cook, W. J., Jeffrey, L. C., Sullivan, M. L., and Vierstra, R. D. (1992). Three-dimensional structure of a ubiquitin-conjugating enzyme (E2). *J Biol Chem* *267*, 15116-15121.
- Cook, W. J., Jeffrey, L. C., Xu, Y., and Chau, V. (1993). Tertiary structures of class I ubiquitin-conjugating enzymes are highly conserved: crystal structure of yeast Ubc4. *Biochemistry* *32*, 13809-13817.
- Cook, W. J., Martin, P. D., Edwards, B. F. P., Yamazaki, R. K., and Chau, V. (1997). Crystal structure of a class I ubiquitin conjugating enzyme (Ubc7) from *Saccharomyces cerevisiae* at 2.9 angstroms resolution. *Biochemistry* *36*, 1621-1627.
- Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* *13*, 289-302.
- de la Torre, J. G., Huertas, M. L., and Carrasco, B. (2000). HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Mag Res* *147*, 138-146.
- 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.
- Dominguez, C., Bonvin, A. M. J. J., Winkler, G. S., van Schaik, F. M. A., Timmers, H. T. M., and Boelens, R. (2004). Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis and docking approaches. *Structure* *12*, 633-644.
- Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J Biomol NMR* *16*, 23-28.
- Duggan, B. M., Legge, G. B., Dyson, H. J., and Wright, P. E. (2001). SANE (Structure Assisted NOE Evaluation): an automated model-based approach for NOE assignment. *J Biomol NMR* *19*, 321-329.
- Engh, R. A., and Huber, R. (1991). Accurate Bond and Angle Parameters for X-Ray Protein-Structure Refinement. *Acta Crystallog A* *47*, 392-400.
- Farrow, N. A., Archer, S. J., Wu, Z. J., Camac, D. M., Parsons, T., Rolfé, M., and Domaille, P. J. (2000). Backbone resonance assignment of human UBC4. *J Biomol NMR* *18*, 363-364.

- Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Formankay, J. D., and Kay, L. E. (1994). Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology-2 Domain Studied by N-15 NMR Relaxation. *Biochemistry* *33*, 5984-6003.
- Giraud, M. F., Desterro, J. M. P., and Naismith, J. H. (1998). Structure of ubiquitin-conjugating enzyme 9 displays significant differences with other ubiquitin-conjugating enzymes which may reflect its specificity for sumo rather than ubiquitin. *Acta Crystallog D* *54*, 891-898.
- Glickman, M. H., and Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev* *82*, 373-428.
- Gonen, H., Bercovich, B., Orian, A., Carrano, A., Takizawa, C., Yamanaka, K., Pagano, M., Iwai, K., and Ciechanover, A. (1999). Identification of the ubiquitin carrier proteins, E2s, involved in signal-induced conjugation and subsequent degradation of IkappaBalpha. *J Biol Chem* *274*, 14823-14830.
- Guntert, P. (2003). Automated NMR protein structure calculation. *Prog Nuc Mag Res Sp* *43*, 105-125.
- Guntert, P., Berndt, K. D., and Wuthrich, K. (1993). The Program Asno for Computer-Supported Collection of Noe Upper Distance Constraints as Input for Protein-Structure Determination. *J Biomol NMR* *3*, 601-606.
- Hamilton, K. S., Ellison, M. J., Barber, K. R., Williams, R. S., Huzil, J. T., McKenna, S., Ptak, C., Glover, M., and Shaw, G. S. (2001). Structure of a conjugating enzyme-ubiquitin thiolester intermediate reveals a novel role for the ubiquitin tail. *Structure (Camb)* *9*, 897-904.
- Hare, B. J., and Wagner, G. (1999). Application of automated NOE assignment to three-dimensional structure refinement of a 28 kDa single-chain T cell receptor. *J Biomol NMR* *15*, 103-113.
- Herrmann, T., Guntert, P., and Wuthrich, 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.
- Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P. M., Huibregtse, J. M., and Pavletich, N. P. (1999). Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* *286*, 1321-1326.
- Jiang, F., and Basavappa, R. (1999). Crystal structure of the cyclin-specific ubiquitin-conjugating enzyme from clam, E2-C, at 2.0 Å resolution. *Biochemistry* *38*, 6471-6478.
- 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.
- Jorgensen, W. L., and Tirado-rives, J. (1988). The OPLS Potential functions for proteins. Energy minimizations for crystals of cyclin peptides and crambin. *J Am Chem Soc* *110*, 1657-1666.
- Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., and Torchia, D. A. (1992). Pulse Sequences for Removal of the Effects of Cross-Correlation between Dipolar and Chemical-Shift Anisotropy Relaxation Mechanism on the Measurement of Heteronuclear T1 and T2 Values in Proteins. *J Mag Res* *97*, 359-375.
- Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* *14*, 51-55, 29-32.
- Korzhev, D. M., Skrynnikov, N. R., Millet, O., Torchia, D. A., and Kay, L. E. (2002). An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J Am Chem Soc* *124*, 10743-10753.

- Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24, 946-950.
- Lin, Y., Hwang, W. C., and Basavappa, R. (2002). Structural and functional analysis of the human mitotic-specific ubiquitin-conjugating enzyme, UbcH10. *J Biol Chem* 277, 21913-21921.
- Linge, J. P., Habeck, M., Rieping, W., and Nilges, M. (2004). Correction of spin diffusion during iterative automated NOE assignment. *J Magn Reson* 167, 334-342.
- Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001). Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol* 339, 71-90.
- Linge, J. P., Williams, M. A., Spronk, C., Bonvin, A., and Nilges, M. (2003). Refinement of protein structures in explicit solvent. *Proteins* 50, 496-506.
- Lipari, G., and Szabo, A. (1982). Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of the Experimental results. *J Am Chem Soc* 104, 4559-4570.
- Liu, Q., Yuan, Y. C., Shen, B. H., Chen, D. J., and Chen, Y. (1999). Conformational flexibility of a ubiquitin conjugation enzyme (E2). *Biochemistry* 38, 1415-1425.
- Loria, J. P., Rance, M., and Palmer, A. G. (1999). A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR spectroscopy. *J Am Chem Soc* 121, 2331-2332.
- Meiboom, S., and Gill, D. (1958). Modified spin-echo method for measuring nuclear relaxation times. *Rev Scient Instr* 29, 688-691.
- Merritt, E. A., and Murphy, M. E. P. (1994). Raster3D version 2.0: A program for photorealistic molecular graphics. *Acta Crystallog D* D50, 869-873.
- Miura, T., Klaus, W., Ross, A., Guntert, P., and Senn, H. (2002). The NMR structure of the class I human ubiquitin-conjugating enzyme 2b. *J Biomol NMR* 22, 89-92.
- Moraes, T. F., Edwards, R. A., McKenna, S., Pastushok, L., Xiao, W., Glover, J. N., and Ellison, M. J. (2001). Crystal structure of the human ubiquitin conjugating enzyme complex, hMms2-hUbc13. *Nat Struct Biol* 8, 669-673.
- Mulder, F. A. A., de Graaf, R. A., Kaptein, R., and Boelens, R. (1998). An off-resonance rotating frame relaxation experiment for the investigation of macromolecular dynamics using adiabatic rotations. *J Mag Res* 131, 351-357.
- Muller, S., Hoegge, C., Pyrowolakis, G., and Jentsch, S. (2001). SUMO, ubiquitin's mysterious cousin. *Nat Rev Mol Cell Biol* 2, 202-210.
- Nilges, M., Macias, M. J., O'Donoghue, S. I., and Oschkinat, H. (1997). Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol* 269, 408-422.
- Palmer, A. G., Rance, M., and Wright, P. E. (1991). Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance C-12 Heteronuclear Nmr-Spectroscopy. *J Am Chem Soc* 113, 4371-4380.
- Pawley, N. H., Wang, C. Y., Koide, S., and Nicholson, L. K. (2001). An improved method for distinguishing between anisotropic tumbling and chemical exchange in analysis of N-15 relaxation parameters. *J Biomol NMR* 20, 149-165.
- Peng, J. W., Thanabal, V., and Wagner, G. (1991). 2d Heteronuclear NMR Measurements of Spin-Lattice Relaxation-Times in the Rotating Frame of X Nuclei in Heteronuclear Hx Spin Systems. *J*

Mag Res 94, 82-100.

Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Ann Rev Biochem* 70, 503-533.

Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234, 779-815.

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nuc Mag Res Sp* 34, 93-158.

Scheffner, M., Huibregtse, J. M., and Howley, P. M. (1994). Identification of a human ubiquitin-conjugating enzyme that mediates the E6-AP-dependent ubiquitination of p53. *Proc Natl Acad Sci U S A* 91, 8797-8801.

Seufert, W., and Jentsch, S. (1990). Ubiquitin-conjugating enzymes UBC4 and UBC5 mediate selective degradation of short-lived and abnormal proteins. *Embo J* 9, 543-550.

Stancovski, I., Gonen, H., Orian, A., Schwartz, A. L., and Ciechanover, A. (1995). Degradation of the proto-oncogene product c-Fos by the ubiquitin proteolytic system in vivo and in vitro: identification and characterization of the conjugating enzymes. *Mol Cell Biol* 15, 7106-7116.

Tjandra, N., Feller, S. E., Pastor, R. W., and Bax, A. (1995). Rotational diffusion anisotropy of human ubiquitin from ¹⁵N NMR relaxation. *J Am Chem Soc* 117, 12562-12566.

Tjandra, N., Garrett, D. S., Gronenborn, A. M., Bax, A., and Clore, G. M. (1997). Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nat Struct Biol* 4, 443-449.

Tong, H., Hateboer, G., Perrakis, A., Bernards, R., and Sixma, T. K. (1997). Crystal structure of murine/human Ubc9 provides insight into the variability of the ubiquitin-conjugating system. *J Biol Chem* 272, 21381-21387.

VanDemark, A. P., Hofmann, R. M., Tsui, C., Pickart, C. M., and Wolberger, C. (2001). Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. *Cell* 105, 711-720.

Weissman, A. M. (2001). Themes and variations on ubiquitylation. *Nat Rev Mol Cell Biol* 2, 169-178.

Worthylake, D. K., Prakash, S., Prakash, L., and Hill, C. P. (1998). Crystal structure of the *Saccharomyces cerevisiae* ubiquitin-conjugating enzyme Rad6 at 2.6 Å resolution. *J Biol Chem* 273, 6271-6276.

Wu, P. Y., Hanlon, M., Eddins, M., Tsui, C., Rogers, R. S., Jensen, J. P., Matunis, M. J., Weissman, A. M., Wolberger, C. P., and Pickart, C. M. (2003). A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *Embo J* 22, 5241-5250.

Zheng, N., Wang, P., Jeffrey, P. D., and Pavletich, N. P. (2000). Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* 102, 533-539.



Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis and docking approaches

NMR and docking results are based on:

C. Dominguez, A. M. J. J. Bonvin, G. S. Winkler, F. M. A. van Schaik, H. Th. M. Timmers, and R. Boelens. Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis and docking approaches. *Structure*, 12, 633-644 (2004)

Site-directed mutagenesis results are based on:

G. S. Winkler, T. K. Albert, C. Dominguez, Y. I. A. Legtenberg, R. Boelens, and H. Th. M. Timmers. An altered-specificity ubiquitin-conjugating enzyme / ubiquitin-protein ligase pair. *J. Mol. Biol.*, 337, 157-165 (2004)

Abstract

The protein CNOT4 possesses an N-terminal RING finger domain that acts as an E3 ubiquitin ligase and specifically interacts with UbcH5B, a ubiquitin conjugating enzyme. The structure of the CNOT4 RING domain has been solved and the amino acids important for the binding to UbcH5B have been mapped. Here, the residues of UbcH5B important for the binding to CNOT4 RING domain were identified by NMR chemical shift perturbation experiments and these data were used to generate structural models of the complex with the program HADDOCK. Based on the chemical shift perturbation mapping and the docking results, mutagenesis experiments were performed that identified several basic residues of UbcH5B important for binding to CNOT4. Importantly, concomitant charge-alteration of Glu49 of CNOT4 and Lys63 of UbcH5B restored binding indicating an electrostatic interaction between these two residues. Together with the NMR data, these additional biochemical data were included in a second docking and comparisons of the resulting model with the structure of the c-Cbl/UbcH7 complex reveal some significant differences, notably at specific residues, and give structural insights into the E2/E3 specificity.

Introduction

The RING finger protein CNOT4 is a component of the CCR4-NOT complex, a global repressor of RNA polymerase II transcription (Collart, 2003; Denis and Chen, 2003). The structure of the RING finger domain of CNOT4, consisting of the 78 N-terminal residues has been solved by NMR (Hanzawa et al., 2001). Recently, CNOT4 has been identified as an E3 ubiquitin-protein ligase, and the RING domain of CNOT4 has been shown to be necessary and sufficient to specifically interact with the ubiquitin conjugating enzyme, UbcH5B (Albert et al., 2002). These results link CCR4-NOT mediated transcription repression with the ubiquitination pathway. Whether this ubiquitination is related to degradation of transcription complexes by the 26S proteasome or other regulatory functions is unclear.

Ubiquitination of a substrate involves three enzymes: a ubiquitin activating enzyme (E1), a ubiquitin conjugating enzyme (E2) and, finally, a ubiquitin protein ligase (E3) (Glickman and Ciechanover, 2002; Pickart, 2001; Weissman, 2001). Different classes of E3 ligases have been found that mediate substrate ubiquitination. E3 enzymes can be distinguished by their E2-interacting domains, which include the HECT and the RING domains (Glickman and Ciechanover, 2002; Pickart, 2001). Therefore, specific E2/E3 pairs are thought to be responsible for mediating target recognition and subsequent ubiquitination. Structures of nine E2 enzymes from different species in the free form have been solved by X-ray crystallography (Cook et al., 1992; Cook et al., 1993; Cook et al., 1997; Hamilton et al., 2001; Jiang and Basavappa, 1999; Lin et al., 2002; Tong et al., 1997; VanDemark et al., 2001; Worthylake et al., 1998) and one was solved by NMR (Miura et al., 2002). The structure of the human E2 enzyme UbcH5B, however, has not been determined. Furthermore, five structures of E2 in complex with various other proteins have been solved by X-ray crystallography (Bernier-Villamor et al.,

2002; Huang et al., 1999; Moraes et al., 2001; VanDemark et al., 2001; Zheng et al., 2000). All structures possess the same fold corresponding to an N-terminal α -helix, followed by a four-stranded anti-parallel β -sheet, and three α -helices. The ternary fold is well conserved among the structures both in free form and in complex with other proteins.

Among the E2 structures in complex, two are of particular interest for the E2/E3 recognition: the X-ray structures of the ubiquitin conjugating enzyme UbcH7 in complex with the E6AP ubiquitin ligase (Huang et al., 1999) and of UbcH7 in complex with the c-Cbl ubiquitin ligase (Zheng et al., 2000). E6AP contains a HECT domain and c-Cbl possesses a C-terminal RING finger domain. Both domains are responsible for the interaction with UbcH7. Besides the RING finger domain of c-Cbl, a second domain, the so-called linker region, also makes contacts with UbcH7. The two structures give considerable insights into the molecular basis of the E2/E3 recognition. Although the RING domain of c-Cbl and the HECT domain of E6AP are structurally unrelated, they both bind UbcH7 in a very similar manner. In both complexes the same residues of UbcH7 are involved in the interaction and these residues are highly conserved within the E2 enzymes suggesting that other E2/E3 complexes could follow a similar arrangement. There are a limited number of E2 enzymes (13 in *Saccharomyces cerevisiae* and more in higher organisms) whereas the number of E3 protein ligases is still increasing (more than 350 RING domains have been identified in human). This suggests that one E2 can recognize different E3s and that one E3 will only recognize one or few E2s. Correspondingly, it has been shown, that the E2/E3 interaction is highly specific: for example the RING domain of CNOT4 interacts functionally with UbcH5B but not with several other E2 enzymes (Winkler et al., 2004). Similarly, it has recently been shown that the RING protein BRCA1 can interact both with UbcH5C and UbcH7 but that only UbcH5C is functional in BRCA1-mediated substrate independent Ub-ligase assay (Brzovic et al., 2003). The c-Cbl/UbcH7 and UbcH7/E6-AP complexes involving the same E2 enzyme are thus not sufficient to understand the molecular basis of the E2/E3 specificity and structural information from complexes involving other ubiquitin conjugating enzymes is crucial to understand the E2/E3 specificity at a molecular level.

It has been previously reported that the CNOT4 RING finger domain interacts with UbcH5B, and the amino acids of CNOT4 important for the binding to UbcH5B have been determined by NMR (Albert et al., 2002). Now, we performed the complementary NMR titration experiments to identify the amino acids of UbcH5B that are involved in the binding to the CNOT4 RING finger domain. Using data from the two titrations we applied a docking approach HADDOCK (Dominguez et al., 2003), that we have recently developed, to generate a structural model of the UbcH5B/CNOT4 RING complex. The initial docking calculations resulted in two possible models of the UbcH5B/CNOT4 complex. Based on the docking results, we mapped by site-directed mutagenesis a number of basic UbcH5B residues important for the interaction with CNOT4 RING domain. We show that reciprocal substitution of charged amino acids of the UbcH5B-CNOT4 interface result in a functional E2/E3 pair. This additional biochemical information was used together with the NMR titration restraints in a second docking calculation resulting in a uniquely defined complex. Our final experimentally based model of the UbcH5B/CNOT4 complex reveals significant differences in terms of

intermolecular contacts as compared to the c-Cbl/UbcH7 complex, especially in terms of hydrogen bonding and salt-bridges. These differences indicate that, although similar and well-conserved regions of different E2 enzymes and E3 ligases are involved in binding, the intermolecular interactions involve different amino acids and different kinds of forces that can explain the observed E2/E3 specificity of the UbcH5B/CNOT4 complex.

Results and Discussion

Assignment of UbcH5B and secondary structure prediction

The assignment of UbcH5B was previously reported (Farrow et al., 2000). However, the data were collected at different temperature, pH and salt concentration than used here. Therefore, 3D TOCSY- $(^1\text{H}, ^{15}\text{N})$ HSQC, 3D NOESY- $(^1\text{H}, ^{15}\text{N})$ HSQC, 3D HNCACB and 3D CBCACONH were recorded to assign the backbone chemical shifts in our conditions (300 K, pH 7.0, 150 mM KCl). Side chain protons were assigned from 3D (H)CCH TOCSY and H(C)CH TOCSY experiments. Talos (Cornilescu et al., 1999) was used to predict the dihedral angles of the protein based on the $\text{C}\alpha$ and the $\text{C}\beta$ chemical shifts. These predictions correlate well with the consensus secondary structure elements found in the known E2 structures (data not shown).

Structural model of UbcH5B

Although the structure of UbcH5B is not available in the protein data bank (PDB), the amino acid sequence of UbcH5B is highly homologous to that of other E2 enzymes. All E2 structures already described are structurally very similar and our secondary structure prediction based on the NMR $\text{C}\alpha$ and $\text{C}\beta$ chemical shifts of UbcH5B correlates well with the E2 consensus structure. We therefore decided to generate a structural model of UbcH5B based on homologous structures. A Blast search of the UbcH5B sequence against the protein data bank identified the most homologous E2 enzymes. In particular, UbcH5B displays 90% homology and 80% identity with yeast Ubc4 (PDB: 1QCQ). Therefore, yeast Ubc4 was used to build a homology model of UbcH5B. The Ramachandran plot of the generated model shows 89.8% of the residues in the most favoured regions, 9.5% in the additional allowed regions, 0.6% in the generously allowed region and 0.2% in the disallowed regions. The backbone RMSD between the model and the yeast Ubc4 structure is 0.20 Å. A 5 nanoseconds molecular dynamic simulation in explicit solvent was performed in order to assess the stability of the model and to obtain an ensemble of starting structures for the docking by extracting PDB files from the trajectory (see Material and Methods). Analysis of these PDB files shows that few structural rearrangements occur during the simulation. The mean RMSD among all structures extracted from the trajectory is 1.2 Å for the backbone atoms (Figure 1) and 1.7 Å for all heavy atoms.

Mapping of the CNOT4-N78 binding site of UbcH5B

The UbcH5B binding site of the CNOT4 RING domain has been previously determined by

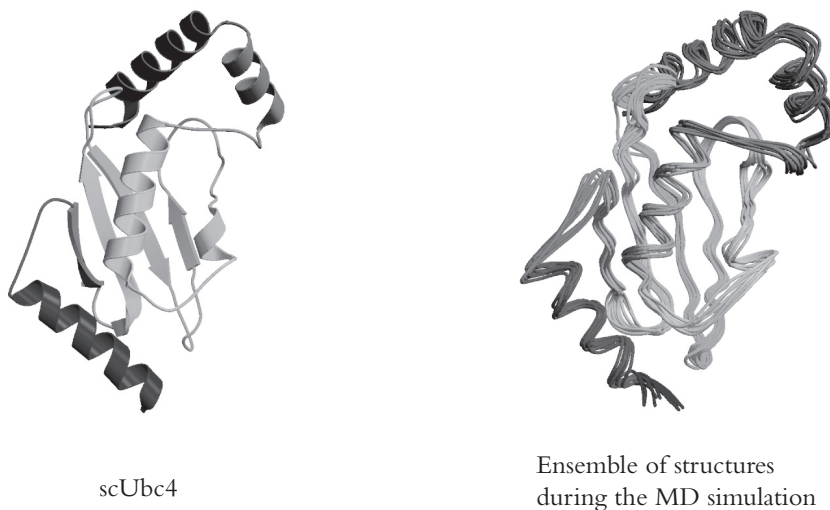


Figure 1: X-ray structure of yeast Ubc4 and ensemble of UbcH5B models during the 5ns molecular dynamic simulation. Figures were generated with the programs Molscript (Kraulis, 1991) and Raster3D (Merrit and Murphy, 1994).

NMR chemical shift perturbations in a titration of ^{15}N -labeled CNOT4-N78 (comprising the 78 N-terminal amino acids of CNOT4) with unlabeled UbcH5B (Albert et al., 2002). The residues of CNOT4-N78 having the highest chemical shift perturbation upon binding to UbcH5B are Leu16, Cys17, Met18, Cys41, Asp48, Glu49 and Arg57. The mapping of these residues onto the CNOT4 RING finger NMR structure (PDB: 1E4U) (Hanzawa et al., 2001) reveals three distinct regions. The first region defines the shallow groove of the α helix (residues 16 to 18) and the two other regions correspond to two zinc chelating loops (residues 48, 49 and 57).

In this study, we carried out the reciprocal experiment using ^{15}N -labeled UbcH5B and unlabeled CNOT4-N78 to identify the residues of UbcH5B involved in the interaction with the CNOT4 RING finger. The NMR titration was performed by increasing the UbcH5B/CNOT4 RING finger molar ratio up to a final 1:2 molar ratio. We analyzed the UbcH5B chemical shift perturbation upon binding to CNOT4-N78 and characterized the amino acids potentially involved in the interaction (Figure 2). The largest combined chemical shift differences (higher than 0.1 ppm) are observed in the first α helix (Ala2, Leu3, Arg5, Ile6, Glu9, Leu10, Asn11, and Asp12), the L1 loop connecting the third and fourth strand of the β -sheet (residue Thr58, Asp59, Tyr60 and Lys63) and the L2 loop connecting the fourth β -strand and the H2 α -helix (residues Ser94, Ala96, Thr98, and Ile99). These residues define three distinct interacting regions of UbcH5B, which are located on one side of the molecule. This is consistent with previous data showing that CNOT4-N78 displays three distinct interacting regions upon binding to UbcH5B, which are localized around a hydrophobic cleft (Albert et al., 2002).

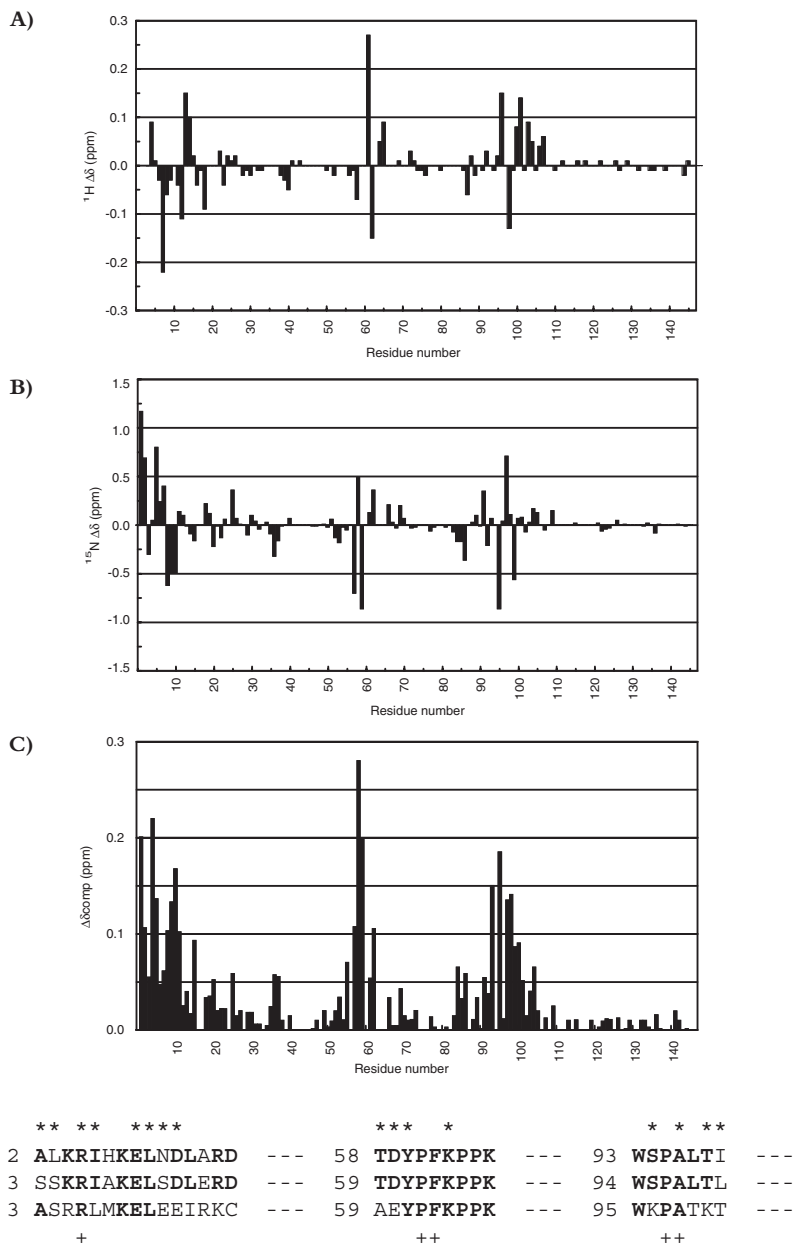


Figure 2: Titration of CNOT4 with UbcH5B. Chemical shifts perturbation of ^{15}N -labeled UbcH5B upon titration of CNOT4-N78 as a function of the UbcH5B amino acid for the **A)** amide proton and **B)** the amide nitrogen. **C)** Combined chemical shift differences ($\Delta\delta = [(\delta_{\text{HN}})^2 + (\delta_{\text{N}}/6.51)^2]^{1/2}$). **D)** Sequence alignment of UbcH5B, yeast UBC4 and UbcH7 encompassing the $\alpha 1$ helix and the L1 and L2 loops. The amino acids of UbcH5B displaying a $\Delta\delta_{\text{comp}} > 0.1 \text{ ppm}$ (*), and the amino acids of UbcH7 involved in the binding with c-Cbl (+) are marked.

Ensemble docking of the UbcH5B/CNOT4 RING finger complex

We used HADDOCK (Dominguez et al., 2003) to generate a structural model of the UbcH5B/CNOT4 RING finger domain based on the NMR titration data. A new feature that we added to HADDOCK consists of the possibility of starting from ensembles of structures instead of a single structure. The structure of a protein free in solution is usually different than when forming a complex with a partner. Even when very small structural rearrangements occur (that can be monitored by NMR or circular dichroism for example) the side chain orientation at the interface may be much more different. In HADDOCK, the proteins are treated as rigid bodies at the first stage of the docking. Since no flexibility is allowed at this stage, no side chains or loop rearrangement can occur which might lead to wrong initial orientations of the complex. Such orientations are generally not corrected during the subsequent HADDOCK semi-flexible molecular dynamic refinement stages. A solution to overcome this problem can be to start from an ensemble of structures instead of a single molecule. An NMR ensemble of structures will reflect flexibility in loops and will show slight differences in side chain orientation especially for solvent exposed residues. The different side chain and/or backbone orientations allow a better sampling of all conformational possibilities during the rigid body docking step. However, also in the case that an ensemble of NMR models or different X-ray structures or homology models are not available, an ensemble can be generated by applying a short molecular dynamics simulation. The different structures can then be extracted from the trajectory of the simulation.

The performance of ensemble docking was tested on the EIIA-HPr complex (Wang et al., 2000) starting from the X-ray structure of EIIA (Worthylake et al., 1991) and the 30 NMR structures of HPr (van Nuland et al., 1994). After the water refinement, only two clusters were obtained using a 1.5 Å cut-off for the ensemble docking instead of 14 when starting from single structures (Dominguez et al., 2003). The model obtained starting from an ensemble of structures is closer to the target complex than the single structure docking solution as can be seen from Table 1. In order to get more insights into the validity of our docking approach, we also analyzed the intermolecular contacts presents in our best models and compared them with the contacts presents in the NMR structure. The analysis was performed using LIGPLOT (Wallace et al., 1995) on the 10 best solutions of the best cluster for the ensemble docking (this cluster contains 87 solutions) and on the 7 solutions of the best cluster for the original docking (see Table 1). The HADDOCK solutions reproduced 100% of the hydrogen bonds and/or salt bridges found in the NMR structure and 63% of the non-bonded contacts (Table 1). Note that for the NMR structure the analysis was based on a single structure, the deposited minimized average, and no contact statistics could thus be obtained. This structure was also calculated in the absence of any attractive non-bonded energy term. Next to those “native” contacts, 5 additional intermolecular hydrogen bonds/salt bridges were detected in the HADDOCK models (Table 1). For the docking of the UbcH5B/CNOT4 RING complex, 30 NMR structures of the CNOT4 RING domain deposited in the PDB (PDB:1E4U) (Hanzawa et al., 2001) and 11 models of UbcH5B obtained during the molecular dynamic simulations (the initial homology model and 10 molecular dynamic structures taken at 0.5 ns intervals) (see Material and Methods) were used.

Table 1: Performance of ensemble docking with HADDOCK for EIIA-HPr complex

	Original protocol	Ensemble docking	NMR structure
Backbone RMSD from NMR structure (Å)	2.5 ± 0.2	1.4 ± 0.2	-
Backbone interface RMSD from NMR structure (Å) ^a	2.1 ± 0.1	1.7 ± 0.2	-
Intermolecular energy (kcal mol ⁻¹) ^b	-392 ± 49	- 443 ± 92	-
Buried surface area (Å ²) ^c	1453 ± 114	1676 ± 150	1365
Intermolecular contacts ^d			
Number of H bonds / % of native H bond	9 / 40%	10 / 100%	5
Number of non bonded contacts / % of native contact	11 / 21%	18 / 63%	19
Total number of intermolecular contacts / % of native contacts	20 / 25%	28 / 71%	24
% of total native contact present in at least one structure	58%	96%	

- a) The interface consists of residues 36 to 48, 66 to 82, 84 to 88, 92 to 98 and 139 to 143 for EIIA and residues 10 to 22 and 46 to 58 for HPr.
- b) The intermolecular energy consists of the sum of the van der Waals, the electrostatic and the AIR energy terms.
- c) Calculated with NACCESS (Hubbard and Thornton, 1993) using a 1.4 Å radius probe.
- d) Intermolecular contacts were analyzed with DIMPLOT (Wallace et al., 1995) (see Material and methods) and are reported if present in at least half of the analyzed structures. The percentages correspond to the percentage of native hydrogen bond or non-bonded contact in the NMR structure that are also present in at least half of the analyzed structures

Based on the NMR titration data, we first selected all residues having a combined chemical shift perturbation upon complex formation higher than 0.1 and 0.05 ppm for UbcH5B and CNOT4, respectively. The cut-off of 0.1 and 0.05 ppm was defined as all residues having a chemical shift perturbation higher than the average chemical shift perturbation along the sequence plus one standard deviation, in combination with a solvent accessibility criteria. In total, 12 and 8 active residues and 7 and 12 passive residues for CNOT4 and UbcH5B, respectively, were used to define 20 ambiguous interaction restraints (AIRs) to drive the docking process (Table 2) (see Material and Methods). After water refinement, two clusters were obtained (Figure 3) containing 61 and 127 structures, respectively. The best 10 structures of these two clusters have average total interaction energies of

-522 ± 17 and -487 ± 29 kcal mol⁻¹, AIR energies of 2.23 ± 2.07 and 1.28 ± 0.72 kcal mol⁻¹ and buried surface areas of 1388 ± 81 and 1523 ± 83 Å², respectively. The second cluster is the most populated, satisfies best the experimental restraints and possesses the largest interface. Its total intermolecular energy is somewhat higher than the first cluster, due mainly to differences in electrostatic energies. In order to differentiate between these two clusters, we analyzed them in details. The two clusters correspond to models with ~ 180 degrees rotation of CNOT4 around the intermolecular axis defined by the two proteins (Figure 3). The difficulty in distinguishing between them is due to the symmetry of the charge distribution at the interface. Two residues of CNOT4 showing a significant chemical shift perturbation upon complex formation are the amino acids Asp48 and Glu49. We thus investigated the hydrogen bond pattern in the two clusters. In the best solution of cluster 1, Asp48 and Glu49 of CNOT4 form salt-bridges with the Lys4 and Lys8 of Ubch5B, while in the best solution of cluster 2, these two residues pair with Lys63 of Ubch5B (Figure 3).

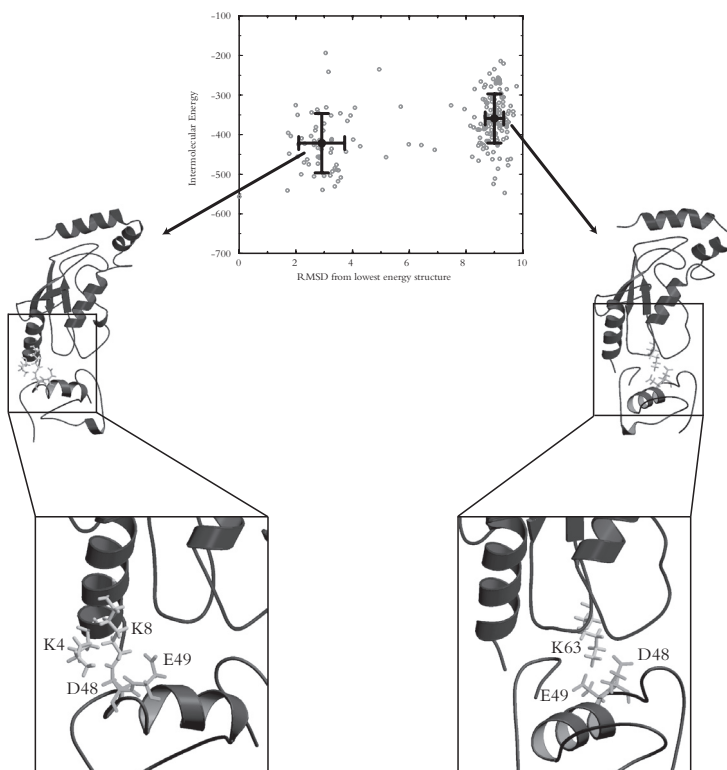


Figure 3: Docking results of the Ubch5B/CNOT4 complex when using NMR perturbation data only. Plot of the intermolecular energy as a function of the RMSD from the lowest energy structure for the 200 calculated solutions. Two clusters are defined. Asp48 and Glu49 of CNOT4 are hydrogen bonded to Lys4 and Lys8 of Ubch5B in the first cluster (left) and to Lys63 of Ubch5B in the second cluster (right).

Table 2: Active and passive residues of UbcH5B and CNOT4-N78

<i>UbcH5B</i>		<i>CNOT4</i>	
Active ^a	Passive ^b	Active ^a	Passive ^b
Leu3	Met1	Leu16	Glu13
Arg5	Lys4	Cys17	Pro15
Thr58	His7	Met18	Pro20
Asp59	Lys8	Glu19	His43
Lys63	Asp29	Phe40	Arg44
Ser94	Met30	Asp48	Thr47
Ala96	Phe31	Glu49	Pro54
Thr98	His55	Asn50	
	Phe62	Leu52	
	Lys66	Ala55	
	Gln92	Cys56	
	Pro95	Arg57	

- a) The active residues correspond to the residues having a significant NMR chemical shift perturbation during the NMR titration experiments and that are high solvent accessible.
- b) The passive residues correspond to all surface neighbors of the active residues that are solvent accessible.

Identification of basic residues essential for UbcH5B/CNOT4 interactions

Consistent with the chemical shift perturbation of Asp48 and Glu49 of CNOT4 in the interaction with UbcH5B, substitution of Glu49 by a lysine (E49K) abolished binding to UbcH5B (Albert et al., 2002). We investigated whether one or several of the basic UbcH5B residues were required for the interaction with CNOT4. Thus, basic amino acids located in the $\alpha 1$ helix and loop L1 were mutated and tested in combination with the CNOT4 RING in a yeast two-hybrid interaction assay. Similar levels of all proteins were present in the transformants as determined by immunoblotting (Figure 4A). Remarkably, alteration of residue Lys4 or Arg5 into either alanine or glutamic acid abolished all binding with the CNOT4 RING (Figure 4A and B). Notably, the substitution Lys8Ala did only partially interfere with CNOT4 binding, although substitution Lys8Glu severely disrupted the interaction, indicating that Lys8 may be involved in an electrostatic interaction. As expected from these results, no interaction was observed when both Lys4 and Lys8 were substituted with acidic residues. Interestingly, substitution of loop L1 residue Lys63 with either alanine or glutamic acid also severely reduced CNOT4 binding, but alterations of residue Lys66 did not (Figure 4A and B). This analysis suggests important roles for basic residues in the $\alpha 1$ helix and loop L1 in the association of UbcH5B and CNOT4. Notably, Arg5 and Lys63 are present in multiple E2 enzymes while the presence of lysine at positions 4, 8, and 66 is less conserved.

Compensatory amino acid substitutions in CNOT4 restore binding to an interaction-deficient UbcH5B mutant

Next, we investigated whether any of the UbcH5B basic residues in the $\alpha 1$ helix and/or loop

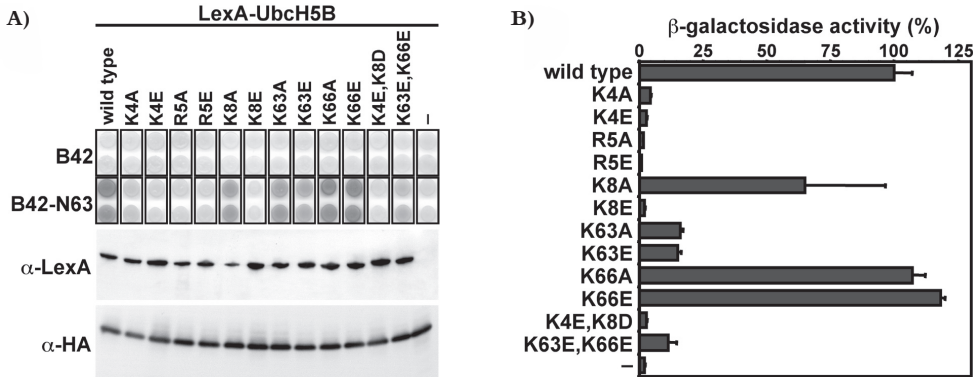


Figure 4: Essential role for basic residues in the interaction between UbcH5B and CNOT4. **A)** Staining of yeast two-hybrid transformants on X-gal plates representing interactions between UbcH5B fused to the LexA DNA-binding domain and the B42 activation domain or B42 fused to CNOT4 N63 as indicated. Shown in the lower panels are immunoblot analyses of LexA-UbcH5B (α -LexA) and B42-CNOT4 N63 (α -HA) proteins. **B)** Quantitative β -galactosidase reporter assay of yeast two-hybrid interactions shown in A). Activities are represented as a percentage of the activity obtained using wild type UbcH5B and CNOT4 N63, which was arbitrarily set to 100%.

L1 interacted directly with the acidic patch on the interaction surface of CNOT4. Therefore, the panel of UbcH5B mutants in which basic amino acids were substituted with acidic residues was tested for binding with the CNOT4 RING domain containing substitutions Asp48Lys and/or Glu49Lys in the yeast two-hybrid assay. In contrast to the single Asp48Lys alteration, which had no significant effect on UbcH5B binding compared to the wild type interaction, substitution Glu49Lys of CNOT4 severely affected binding with wild type UbcH5B. This was even more pronounced when the substitution Glu48Lys was additionally present (Figure 5A and B). Remarkably, the interaction was restored when specific reciprocal amino acid substitutions were introduced in UbcH5B. Thus, when the Lys63Glu substitution in loop L1 was present or when both Lys63Glu and Lys66Glu were concomitantly altered, the interaction was comparable with the interaction between the wild type proteins. This effect was specific, because alterations in the UbcH5B N-terminal α -helix of the solvent exposed residues Lys4Glu and/or Lys8Glu did not result in restoration of the interaction with any of the CNOT4 mutants (Figure 5A and B). In all experiments, similar amounts of the LexA-UbcH5B variants and the B42-CNOT4 N63 partners were present (Figure 5A). These results indicate that CNOT4 Glu49 and UbcH5B Lys63 are directly interacting. Although Asp48 of CNOT4 does not seem a primary determinant, we note that the interaction between UbcH5B with CNOT4^{Asp48Lys,Glu49Lys} is lower compared to the single mutant CNOT4^{Glu49Lys}. This suggests that residue Asp48 can partially compensate for the absence of Glu49. Additionally, interactions involving CNOT4^{Glu49Lys} and CNOT4^{Asp48Lys,Glu49Lys} with UbcH5B^{Lys63Glu,Lys66Glu} seem to be restored to a slightly higher level compared to that obtained with UbcH5B^{Lys63Glu}. Thus, substitution of Lys66Glu generates an additional charge contribution to UbcH5B. The important conclusion of these experiments,

however, is that a single conserved lysine residue located in loop L1 of UbcH5B and an acidic residue on the interaction surface of the CNOT4 RING domain can dominantly influence the interaction between an E2 enzyme and a RING domain E3 ligase.

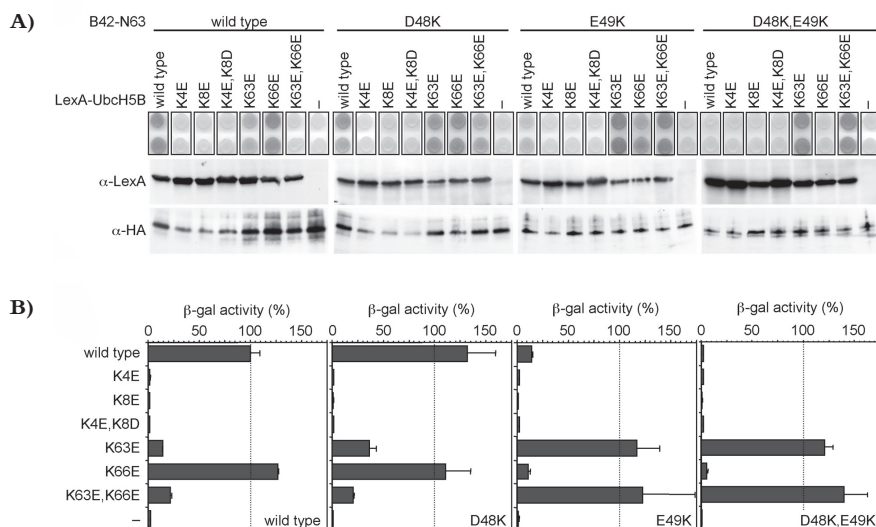


Figure 5: Compensatory mutations in CNOT4 restore binding with UbcH5B containing specific amino acid substitutions. **A)** Staining of yeast two-hybrid transformants on X-gal plates representing interactions between LexA-UbcH5B proteins and B42-CNOT4 N63 proteins as indicated. Shown in the lower panels are immunoblot analyses of LexA-UbcH5B (α -LexA) and B42-CNOT4 N63 (α -HA) proteins. **B)** Quantitative β -galactosidase reporter assay of yeast two-hybrid interactions shown in A). Activities are represented as a percentage of the activity obtained using wild type UbcH5B and CNOT4 N63, which was arbitrarily set to 100%.

Ensemble docking based on NMR chemical shift perturbation and mutagenesis data

We performed a new docking by defining, in addition to the previous ambiguous restraints derived from NMR chemical shift perturbation experiments, a restraint between any atom of Glu49 of CNOT4 and any atom of Lys63 of UbcH5B. Now, only one cluster could be detected containing 168 structures out of the 200 calculated when using a 2.5 Å RMSD (199 structures belong to this cluster when using a 3.0 Å RMSD) (Figure 6A). The complex is well defined with an average energy of the 10 best structures of -515 ± 65 kcal mol⁻¹, and an average backbone RMSD of 1.70 ± 0.56 Å at the interface (Table 3). The best solution of the cluster in terms of intermolecular energy (-653 kcal mol⁻¹) is displayed in Figure 6B and is very similar to the best structure of the second cluster in the previous docking (total backbone RMSD of 1.2 Å while the total backbone RMSD between this model and the best model of the first cluster is 9.5 Å). In this cluster, salt-bridges are found between Lys63

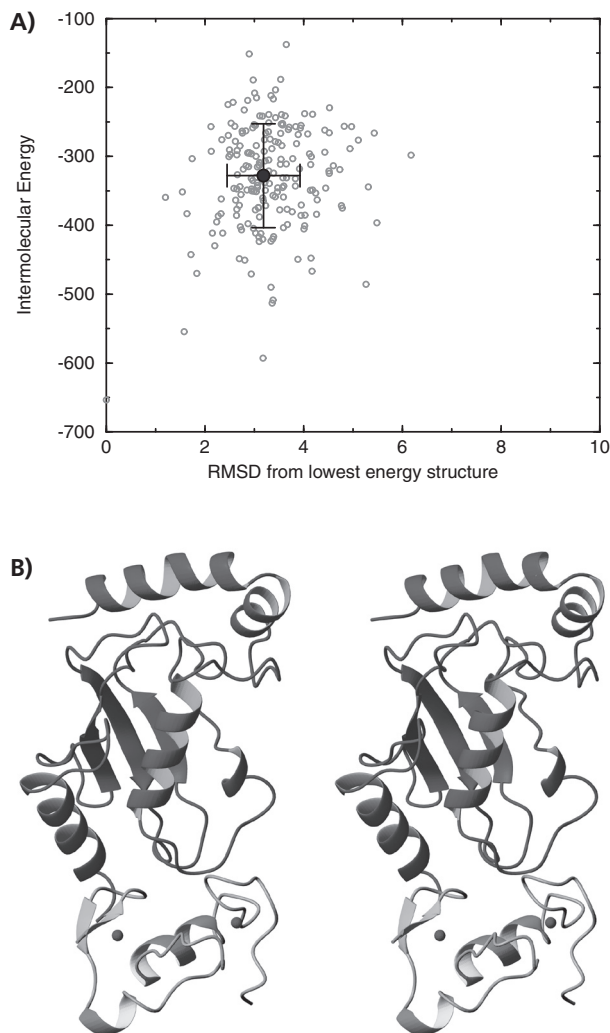


Figure 6: Docking results of the UbcH5B/CNOT4 complex after incorporation of the mutagenesis data. **A)** Plot of the intermolecular energy as a function of the RMSD from the lowest energy structure for the 200 calculated solutions. **B)** Stereo view of the best solution generated by HADDOCK after incorporation of mutagenesis data. This Figure has been generated with the program MOLMOL (Koradi et al., 1996).

of UbcH5B with Asp48 and Glu49 of CNOT4, but also between Lys4 and Lys8 of UbcH5B with Glu13 of CNOT4, and Asp59 of UbcH5B with Arg44 of CNOT4 (Table 4). In addition to this salt-bridges network, a number of van der Waals contacts occur at the interface between residues 1, 4, 5, and 8 of UbcH5B corresponding to the H1 helix and residues 15, 17, 18, 40, and 41 of CNOT4, between residues 62 of the L1 loop of UbcH5B and residues 44, and 45 of CNOT4 and finally between residues 92, 94, 95, and 96 of the L2 loop of UbcH5B and residues 49, 54, and 57 of CNOT4 (Table 4). Concerning the H1 helix, the first residue is not visible in the HSQC spectrum, residue 5 shows a significant perturbation in the NMR titration experiment and residues 4 and 8 have been shown to be involved in

the interaction by mutagenesis experiments. Concerning the loop L1, in addition to the salt bridges involving Asp59 and Lys63 of UbcH5B, both showing a significant perturbation, Phe62 is in van der Waals contact with residues of CNOT4. However, Phe62 does not show a significant perturbation. This could be due to the fact that all contacts between Phe62 and residues of CNOT4 involve the side chain aromatic ring and the perturbation is detected on the amide backbone atom. Finally, concerning loop L2, residues 94 and 96 show significant perturbation while Pro95 is not observed. On the CNOT4 side, all the residues involved in the van der Waals contacts show significant perturbation except for Arg44. The absence of perturbation for Arg44 is striking since this residue in our structural model is salt-bridged to Asp59 of UbcH5B. The NMR titration data also showed that significant perturbation of the

Table 3: Structural statistics of the 10 best UbcH5B/CNOT4 model structures^a

Backbone r.m.s.d. (Å) with respect to mean		
	Flexible interface backbone ^b	1.70 ± 0.56
	All Backbone	2.69 ± 1.46
Number of ambiguous interaction restraints (AIRs)		
	From UbcH5B	12
	From CNOT4	8
	Total AIRs	20
CNS intermolecular energies after water refinement ^c		
	E_{vdw} (kcal mol ⁻¹)	-54 ± 14
	E_{elec} (kcal mol ⁻¹)	-461 ± 75
Buried surface area (Å ²) ^d		1551 ± 117
R.m.s.d from idealized covalent geometry		
	Bonds (Å)	0.003 ± 0.00
	Angles (°)	0.43 ± 0.01
	Impropers (°)	0.36 ± 0.01
Ramachandran analysis		
	Residues in the favoured region (%)	81.1 ± 2.4
	Residues in additional allowed regions (%)	16.8 ± 2.8
	Residues in generously allowed regions (%)	1.3 ± 1.1
	Residues in disallowed regions (%)	0.7 ± 0.3

- a) Based on 10 structures, obtained after flexible docking with HADDOCK followed by refinement in explicit water using ambiguous interaction restraints derived from chemical shift perturbation data (see Material and Methods).
- b) The flexible interface comprises segments M1-P18, V26-W33, T53-A68 and R90-L103 for UbcH5B and V12-E22 and C38-P59 for CNOT4.
- c) The non-bonded energies were calculated with the OPLS parameters (Jorgensen and Tirado-rives, 1988) using a 8.5 Å cut-off.
- d) The buried surface area was calculated with CNS (Brunger et al., 1998) using a 1.4 Å radius water probe and 0.05Å accuracy.

H1 helix of UbcH5B occurred up to residues Asp12. The structural model shows, however, that the last residue of this helix contacting CNOT4 is Lys8. This is consistent with the maximum length of the interacting area defined by the active and passive residues and by the shape of the molecules. The longest distance found in the interacting area of CNOT4 is between Glu19 and Arg57 and is about 27 Å. This distance fits the distance between Lys63 of UbcH5B and Lys8 in the H1 helix. Therefore, we assume that the perturbation observed for residues Glu9, Leu10, Asn11, and Asp12 of UbcH5B are due to an indirect effect of the binding to CNOT4.

Table 4: Intermolecular contacts^a statistics calculated over the ensemble of the ten best structures of cluster 1 for the second HADDOCK run

Interacting residues		Hydrogen bonds			Non bonded contacts	occurrence
UbcH5B	CNOT4	M-M	S-S	M-S		
M1	C17	0	0	0	1	5
M1	P40	0	0	0	1	5
M1	C41	0	0	0	1	6
K4	E13	0	1	0	0	5
K4	M18	0	0	0	1	7
R5	P15	0	0	1	0	6
R5	M18	0	0	0	1	8
K8	E13	0	1	0	0	7
K8	M18	0	0	0	1	7
D59	R44	0	1	0	1	9
P62	R44	0	0	0	1	9
P62	I45	0	0	0	1	9
K63	D48	0	1	0	0	8
K63	E49	0	1	0	0	10
Q92	E49	0	1	0	0	5
Q92	R57	0	0	1	1	6
S94	P54	0	0	1	1	5
S94	R57	0	1	0	1	5
P95	E49	0	0	0	1	5
P95	P54	0	0	0	1	5
A96	P54	0	0	0	1	8

a) Intermolecular contacts were analyzed with DIMPLOT (Wallace et al., 1995) (see Material and methods) and are reported if present in at least 5 of the 10 best structures. The occurrence of main chain-main chain (M-M), side chain-side chain (S-S) and side chain-main chain (S-M) hydrogen bonds is reported. The number of occurrences of a given interaction over the ensemble of 10 best structures is reported in the Table.

Towards the molecular basis of the E2/E3 specificity

Finally, we compared our structural model of the UbcH5B/CNOT4 complex with the X-ray structure of c-Cbl/UbcH7 (Zheng et al., 2000). The two complexes are structurally homologous with a total backbone RMSD of 3.1 Å (UbcH5B/CNOT4 versus c-Cbl/

UbcH7) and the orientation of the RING domain with respect to the E2 enzyme is similar in both complexes (Figure 7A). It has been shown, however, that the CNOT4 RING finger interacts specifically with UbcH5B and not with UbcH7 despite the fact that in both complexes, the same regions of the E2s are involved in the interaction with the RING. This suggests that although the three regions of UbcH5B and UbcH7 involved in the binding (helix H1, loops L1 and L2) are similar (Figure 2D), the binding properties must be different. It has already been reported that the L2 loop of E2 enzymes plays a role in the E2/E3 specificity (Martinez-Noel et al., 2001). In the UbcH7/c-Cbl crystal structure, the highly conserved Pro96 and Ala97 residues are involved in the binding. When comparing the residues of the L2 loop in UbcH5B and UbcH7, however, it can be noted that the other residues are not conserved between the two E2 enzymes: of these, Ser94, Leu97, Thr98 and Ile99 of UbcH5B, which show significant NMR chemical shift perturbations, are replaced by two lysines (Lys96 and Lys100) and two threonines (Thr99 and Thr101) in UbcH7 (Figure 2D). In our model, we also observe the contacts involving the conserved Pro95 and Ala96. In addition, Ser94 makes a hydrogen bond with Arg57 of CNOT4 and a van der Waals contact with Pro54 (table 4). No direct contact involving Thr98 or Ile99, however, could be detected in our model and, thus we explain the observed chemical shifts to an indirect effect of the complex formation. None of these non-conserved residues in UbcH7 are involved in direct contact with c-Cbl in the crystal structure. These results are in agreement with mutagenesis experiments in which a triple mutant of ubcM4 (Lys96Ser, Thr99Leu, Lys100Thr), that is highly homologous to UbcH7, was able to change the specificity from UbcM4 to UbcH5, demonstrating that one or more of these residues are important for the specificity (Martinez-Noel et al., 2001).

A more detailed comparison of the UbcH5B/CNOT4 structural model with the X-ray structure of the c-Cbl/UbcH7 complex gives further insights into how the E2/E3 specificity occurs. Our NMR titration experiments suggest that the first α -helix of UbcH5B makes many direct contacts with the CNOT4 RING domain whereas in the c-Cbl/UbcH7 complex, this helix provides only a minor contact with the c-Cbl RING finger domain through Arg5 and mainly interacts with the linker region located just N-terminal of the RING domain of c-Cbl (Zheng et al., 2000). Many residues of the first helix of UbcH5B show a significant chemical shift perturbation upon complex formation and mutagenesis data showed that Lys4 and to a lesser extent Lys8 are critical residues for the binding to CNOT4-N78. The CNOT4 RING domain is located at the N-terminus of the protein and does not possess this linker region. As a control, we performed a titration of ^{15}N -UbcH5B with unlabeled CNOT4-N227 (the 227 N-terminal amino acids) containing the RRM (RNA recognition motif) domain to investigate whether another region of CNOT4 is involved in the interaction with the UbcH5B H1 helix. However, we did not observe any additional perturbation in the HSQC spectra (data not shown) indicating that, contrary to the c-Cbl/UbcH7 complex, CNOT4 interacts with UbcH5B only through its RING finger domain, and that the H1 helix of UbcH5B is interacting with the CNOT4 RING finger and not with another region of the protein. This is further supported by the fact that the structural model of the UbcH5B/CNOT4 complex shows several contacts between the H1 helix of UbcH5B and the RING domain of CNOT4 that are not present in

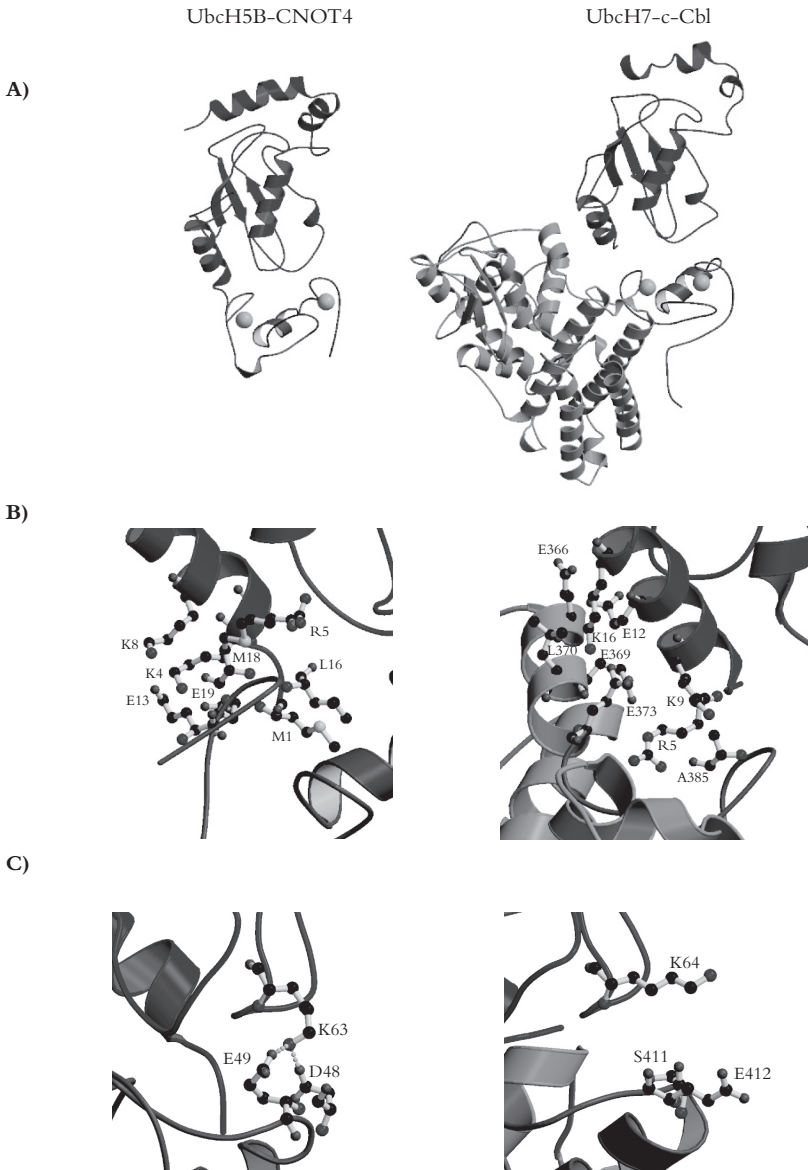


Figure 7: Comparison between the UbcH5B/CNOT4 docking model and the c-Cbl/UbcH7 crystal structure (Zheng et al., 2000). **A)** The orientation of the RING domain compared to the E2 enzyme is similar in both complexes. **B)** The helix $\alpha 1$ of UbcH5B makes many contacts with the CNOT4 RING domain, whereas the helix $\alpha 1$ of UbcH7 interact mainly with the linker region of c-Cbl. Residues of UbcH5B, UbcH7, CNOT4 and c-Cbl that are involved in the interaction are labeled. **C)** In the L1 loop of UbcH5B, residue Lys63 interacts with CNOT4 Asp48 and Glu49 whereas the corresponding and conserved residue of UbcH7 (Lys64) is not in contact with c-Cbl. Figures have been generated with the programs Molscript (Kraulis, 1991) and Raster3D (Merrit and Murphy, 1994).

the c-Cbl/UbcH7 complex (Figure 7B). Furthermore, the contacts that govern the binding of UbcH7 to c-Cbl involve mainly hydrophobic or uncharged residues making multiple van der Waals contacts. In the case of the UbcH5B/CNOT4 complex, however, many charged residues could be detected in the NMR titration experiments suggesting that hydrogen bonds, salt-bridges and thus electrostatic interactions are important in the binding. In the c-Cbl/UbcH7 complex, indeed, residues Pro62 and Phe63 of the L1 loop make close contact with the c-Cbl RING domain. In contrast, our titration of UbcH5B shows significant chemical shift perturbations for Thr58, Asp59, Tyr60, and Lys63, while Phe62 is not affected and Pro61 cannot be observed. The presence of salt bridges in the UbcH5B/CNOT4 complex is further supported by the results of site-directed mutagenesis. Though the residues Lys63 of UbcH5B and Glu49 of CNOT4 are conserved in UbcH7 (Lys64) and c-Cbl (Glu412), they do not form a contact in the crystal structure (Figure 7C).

Taken together, these results give insights into how two different E2/E3 complexes can interact in two distinct ways although their interaction regions are similar with rather high sequence homology. Our present work provides new insight into the origin of the specificity within a given E2/E3 complex in that it indicates that electrostatic interactions are of crucial importance. Other E2/E3 complexes will, however, need to be studied to fully understand the preference of ubiquitin-protein ligases for specific ubiquitin conjugating enzymes.

Material and methods

Recombinant proteins expression, purification and mutagenesis

The bacterial expression plasmids for UbcH5B, CNOT4-N78, and CNOT4-N227 have been described (Albert et al., 2002; Hanzawa et al., 2001). Overexpression of ^{15}N and $^{15}\text{N}/^{13}\text{C}$ -labeled GST-UbcH5B was accomplished by growing *Escherichia coli* BL21(DE3) containing the pGEX2T-UbcH5B plasmid in a synthetic medium (6.0 g/liter $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, 3.0 g/liter KH_2PO_4 , 0.5 g/liter NaCl, 1 mM MgSO_4 , 20 μM CaCl_2 , 36 nM $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 20 μM ZnCl_2 , and 5 mg/liter Thiamine) containing 0.5 g/liter $^{15}\text{NH}_4\text{Cl}$ as the sole nitrogen source and either 4.0 g/liter ^{12}C glucose or 2.0 g/liter ^{13}C glucose as the only carbon source. Induction took place at $A_{600} = 0.7\text{--}0.75$ by addition of 0.4 mM isopropyl- β -thiogalactopyranoside (IPTG) and bacteria were grown 3 h at 30°C and lysed in buffer LB (300 mM KCl, 50 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.1% Triton X-100, 20% sucrose) containing 1 mM dithiothreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride (PMSF), protease inhibitor cocktail (Roche) and 250 $\mu\text{g}/\text{ml}$ lysosyme. After freeze-thawing and sonication, lysates were centrifuged at 25 K rpm in a SW40 rotor for 90 minutes at 4°C. Supernatant was bound to glutathione-agarose in LB buffer without sucrose, washed with buffer WB (150 mM KCl, 20 mM KPi pH 7.0, 10 μM ZnCl_2 , protease inhibitors) and eluted with buffer WB plus 10mM glutathione. The GST tag was cleaved by addition of 3 U of thrombin (Sigma) per mg protein at 37°C for 3 h. Thrombin was inactivated by addition of 0.5 mM PMSF (final concentration) and, subsequently, removed by batch binding to benzamidine-Sepharose6B (Amersham). The UbcH5B sample was concentrated

using a stirred Amicon ultrafiltration cell (Millipore) and further purified by gel filtration chromatography using a Superdex75 16/60 column (Amersham) using WB buffer. Fractions containing UbcH5B were concentrated to a final concentration of 1 mM.

The expression and purification of unlabeled CNOT4-N78 and CNOT4-N227 were as described (Hanzawa et al., 2001). The samples were then dialyzed against NMR buffer (150 mM KCl, 20 mM KPi pH 7.0, 10 μ M ZnCl₂) and concentrated by ultrafiltration to a final concentration of 1 mM.

Site-directed mutagenesis of UbcH5B was carried out using plasmid pEG202-UbcH5B and standard overlap PCR procedures. Mutations in CNOT4 were also generated using PCR techniques with plasmids pJG4-5-N63 and pHis₆-CNOT4 as templates (Albert et al., 2002; Hanzawa et al., 2001). All plasmid constructs were verified by DNA sequencing.

Yeast two-hybrid interaction assay

Growth and manipulation of yeast strain EGY48 was carried out as described (Albert et al., 2000). Proteins were extracted for immunoblotting from yeast transformants using the method of Kushnirov (Kushnirov, 2000). For quantitative determination of β -galactosidase activities, yeast transformants were disrupted with zirconia/silica beads using a mini beadbeater (Biospec Products, 1 min maximum speed at room temperature) in buffer containing 100 mM Tris-HCl pH 8.0, 20% (v/v) glycerol, 1 mM β -mercaptoethanol, 0.5% sodium dodecyl sulphate, and protease inhibitors. After removal of insoluble material by centrifugation, β -galactosidase activities were determined using the Galacto-Light Plus chemiluminescent reporter assay (Tropix) essentially according to the manufacturer's instructions and normalised to total protein content as determined by a Bradford protein assay (Biorad).

UbcH5B modeling

UbcH5B displays 90% homology and 80% identity with the yeast Ubc4 (PDB: 1QCQ) (Cook et al., 1993). We used Modeller4 (Sali and Blundell, 1993) to generate 10 structural models of UbcH5B based on the structure of yeast Ubc4.

Molecular dynamic simulation of UbcH5B

We selected the best model based on Procheck (Laskowski et al., 1993) analysis and applied a 5 nanoseconds molecular dynamic simulation in explicit solvent. PDB files were extracted from the trajectory every 0.5 ns so that, at the end of the simulation, 11 PDB files (the original + 10 from the molecular dynamic) were available as starting structures for the docking. The molecular dynamic simulations were run with the GROMACS3.0 molecular dynamic package (Lindahl et al., 2001) using the GROMOS96 force field (Scott et al., 1999). The structure was solvated in a cubic box of SPC water (Berendsen et al., 1981) using a minimum distance of 14 Å between the protein and the box edges. After a first steepest descent energy minimization with positional restraints on the solute, one chloride counter ion (Cl⁻) was introduced to obtain an electro-neutralized system. A second energy minimization was performed, followed by five successive 20 ps molecular dynamic equilibration runs. During these, the position restraints force constant on the solute Kposre was decreased progressively (1000,1000, 100,

10, 0 kJ mol⁻¹ nm⁻²). A 5 ns production run was then performed at constant temperature (300K) and pressure (1 atm) with weak coupling (0.1 and 1ps⁻¹) to reference T and P baths (Berendsen et al., 1984) using a 4 fs time step for the integration of the equations of motion. Non-bonded interactions were calculated using twin range cut-offs of 0.8 and 1.4 nm. Long range electrostatic interactions beyond the cut-off were treated with the generalized reaction field model using a dielectric constant of 54 (Tironi et al., 1995). The LINCS algorithm (Hess et al., 1997) was used for bond length constraining in conjunction with dummy atoms for the aromatic rings and amino group in side chains (Feenstra et al., 1999) allowing the use of the longer integration time step of 4 fs.

NMR measurements

NMR experiments have been carried out at 300K and pH 7.0 on a Bruker AVANCE600 and AVANCE700 equipped with a triple-resonance z-gradient probe.

For the backbone assignment of UbcH5B, 2D (¹⁵N-¹H)-HSQC, 2D (¹³C-¹H)-HSQC, 3D HNCOC, 3D HNCACB, and 3D CBCA(CO)NH were recorded. The side chain assignment was performed using 3D TOCSY-(¹⁵N-¹H)-HSQC, 3D H(C)CH-TOCSY and 3D (H)CCH-TOCSY spectra (Sattler et al., 1999).

Concerning the chemical shift perturbation experiments, (¹⁵N-¹H)-HSQC spectra were recorded on ¹⁵N-UbcH5B alone and in complex with different ratios of CNOT4-N78 (1:1/8 to 1:2). For all (¹⁵N-¹H)-HSQC spectra, 2048 points with a spectral width of 8012 Hz in the direct dimension and 512 points with a spectral width of 2200 Hz in the indirect dimension were recorded. The number of scans varied between 16 and 64 during the titration and the relaxation delays were set to 1 s. All NMR spectra were processed using the NMRPipe package (Delaglio et al., 1995) and analyzed using the program NMRView (Johnson and Blevins, 1994).

Docking

Docking of the UbcH5B/CNOT4 complex was performed using the software HADDOCK1.1 (Dominguez et al., 2003) in combination with CNS (Brunger et al., 1998) based on the chemical shift perturbation data observed for UbcH5B and for CNOT4 (Albert et al., 2002) upon complex formation. The starting structures for the docking were the 30 structures of CNOT4 deposited in the PDB (PDB:1E4U) (Hanzawa et al., 2001) and the 11 models of UbcH5B generated during the molecular dynamic simulation. For the first docking, the active and passive residues defined for HADDOCK were chosen based on the chemical shift perturbation data and solvent accessibility (Table 2). We first selected all the residues having a combined chemical shift perturbation upon complex formation higher than 0.1 and 0.05 ppm for UbcH5B and CNOT4, respectively. We calculated the solvent accessibility using the program NACCESS (Hubbard and Thornton, 1993) over the ensemble of structures and selected as active residues all the amino acids showing an average relative solvent accessibility (plus standard deviation) higher than 50%. We then selected all surface neighbors amino acids having a high solvent accessibility (>50%) as passive residues. A 2 Å distance was used to define the Ambiguous Interaction Restraints (AIR). Residues 1-18, 26-33, 53-68, 90-103 of UbcH5B

and 12-22, 38-59 of CNOT4 were defined as flexible. In the second docking, an additional restraint between any atoms of Lys63 of Ubch5B and any atoms of Glu49 of CNOT4 was defined based on mutagenesis experiments (Winkler et al., 2004). During the rigid body energy minimization, 1320 structures were calculated (4 calculations for each combination of starting structures). For each of the 1320 combinations, 10 rigid body docking trials were performed and only the solution with lowest energy was kept amounting to a total of 13200 rigid body minimization trials. The 200 best solutions based on the intermolecular energy were used for the semi-flexible simulated annealing followed by a refinement in explicit water. finally, the solutions were clustered using a 2.5 Å RMSD based on the pairwise backbone RMSD matrix after superposition on the backbone of Ubch5B.

Analysis of the intermolecular contacts

Intermolecular contacts (hydrogen bonds and non-bonded contacts) were analyzed with DIMPLOT which is part of the LIGPLOT software (Wallace et al., 1995) using the default settings (3.9 Å heavy-atoms distance cut-off for non-bonded contacts; 2.7 Å and 3.35 Å proton-acceptor and donor-acceptor distance cut-offs respectively with minimum 90° angles (D-H-A, H-A-AA, D-A-AA) for hydrogen bonds (McDonald and Thornton, 1994).

Coordinates

The coordinates of the five best structural models of the Ubch5B/CNOT4 complex have been deposited in the Protein Data Bank (accession code 1UR6) together with the AIR restraints.

Acknowledgements

This investigation was supported by the Netherland Foundation for Chemical Research (NWO-CW) and the Netherland Organization for Scientific Research (NWO). This work was also supported by NWO “Jonge Chemici” and “Pionier” grants to A. M. J. J. Bonvin, and H. Th. M. Timmers, respectively. C. Dominguez was supported financially by the Center for Biomedical Genetics. The authors are grateful to Sandrine Jayne for critical reading of the manuscript.

References

- Albert, T. K., Hanzawa, H., Legtenberg, Y. I. A., de Ruwe, M. J., van den Heuvel, F. A. J., Collart, M. A., Boelens, R., and Timmers, H. T. M. (2002). Identification of a ubiquitin-protein ligase subunit within the CCR4-NOT transcription repressor complex. *Embo J* 21, 355-364.
- Albert, T. K., Lemaire, M., van Berkum, N. L., Gentz, R., Collart, M. A., and Timmers, H. T. (2000). Isolation and characterization of human orthologs of yeast CCR4-NOT complex subunits. *Nucleic Acids Res* 28, 809-817.

- Berendsen, H. C., Postma, J. P. N., van Gunsteren, W. F., and Hermans, J. (1981). Interaction models for water in relation to protein hydration. In *Intermolecular forces*, B. Pullman, ed. (Dordrecht, Reidel), pp. 331-342.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J Chem Phys* *81*, 3684-3690.
- Bernier-Villamor, V., Sampson, D. A., Matunis, M. J., and Lima, C. D. (2002). Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* *108*, 345-356.
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* *54 (Pt 5)*, 905-921.
- Brzovic, P. S., Keeffe, J. R., Nishikawa, H., Miyamoto, K., Fox, D., 3rd, Fukuda, M., Ohta, T., and Klevit, R. (2003). Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proc Natl Acad Sci U S A* *100*, 5646-5651.
- Collart, M. A. (2003). Global control of gene expression in yeast by the Ccr4-Not complex. *Gene* *313*, 1-16.
- Cook, W. J., Jeffrey, L. C., Sullivan, M. L., and Vierstra, R. D. (1992). Three-dimensional structure of a ubiquitin-conjugating enzyme (E2). *J Biol Chem* *267*, 15116-15121.
- Cook, W. J., Jeffrey, L. C., Xu, Y., and Chau, V. (1993). Tertiary structures of class I ubiquitin-conjugating enzymes are highly conserved: crystal structure of yeast Ubc4. *Biochemistry* *32*, 13809-13817.
- Cook, W. J., Martin, P. D., Edwards, B. F. P., Yamazaki, R. K., and Chau, V. (1997). Crystal structure of a class I ubiquitin conjugating enzyme (Ubc7) from *Saccharomyces cerevisiae* at 2.9 angstroms resolution. *Biochemistry* *36*, 1621-1627.
- Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomolecular NMR* *13*, 289-302.
- 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.
- Denis, C. L., and Chen, J. (2003). The CCR4-NOT complex plays diverse roles in mRNA metabolism. *Prog Nucleic Acid Res Mol Biol* *73*, 221-250.
- Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* *125*, 1731-1737.
- Farrow, N. A., Archer, S. J., Wu, Z. J., Camac, D. M., Parsons, T., Rolfé, M., and Dommelle, P. J. (2000). Backbone resonance assignment of human UBC4. *J Biomol NMR* *18*, 363-364.
- Feenstra, K. A., Hess, B., and Berendsen, H. J. C. (1999). Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comput Chem* *20*, 786-798.
- Glickman, M. H., and Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev* *82*, 373-428.
- Hamilton, K. S., Ellison, M. J., Barber, K. R., Williams, R. S., Huzil, J. T., McKenna, S., Ptak, C., Glover, M., and Shaw, G. S. (2001). Structure of a conjugating enzyme-ubiquitin thiolester intermediate reveals a novel role for the ubiquitin tail. *Structure (Camb)* *9*, 897-904.
- Hanzawa, H., de Ruwe, M. J., Albert, T. K., van Der Vliet, P. C., Timmers, H. T. M., and Boelens,

- R. (2001). The structure of the C4C4 ring finger of human NOT4 reveals features distinct from those of C3HC4 RING fingers. *J Biol Chem* 276, 10185-10190.
- Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. (1997). LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18, 1463-1472.
- Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P. M., Huibregtse, J. M., and Pavletich, N. P. (1999). Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* 286, 1321-1326.
- Hubbard, S. J., and Thornton, J. M. (1993). NACCESS (University College London, Department of Biochemistry and Molecular Biology).
- Jiang, F., and Basavappa, R. (1999). Crystal structure of the cyclin-specific ubiquitin-conjugating enzyme from clam, E2-C, at 2.0 Å resolution. *Biochemistry* 38, 6471-6478.
- 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.
- Jorgensen, W. L., and Tirado-rives, J. (1988). The OPLS Potential functions for proteins. Energy minimizations for crystals of cyclin peptides and crambin. *J Am Chem Soc* 110, 1657-1666.
- Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 14, 51-55, 29-32.
- Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24, 946-950.
- Kushnirov, V. V. (2000). Rapid and reliable protein extraction from yeast. *Yeast* 16, 857-860.
- Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26, 283-291.
- Lin, Y., Hwang, W. C., and Basavappa, R. (2002). Structural and functional analysis of the human mitotic-specific ubiquitin-conjugating enzyme, UbcH10. *J Biol Chem* 277, 21913-21921.
- Lindahl, E., Hess, B., and van der Spoel, D. (2001). GROMACS3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 7, 306-317.
- Martinez-Noel, G., Muller, U., and Harbers, K. (2001). Identification of molecular determinants required for interaction of ubiquitin-conjugating enzymes and RING finger proteins. *Eur J Biochem* 268, 5912-5919.
- McDonald, I. K., and Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J Mol Biol* 238, 777-793.
- Merrit, E. A., and Murphy, M. E. P. (1994). Raster3D version 2.0: A program for photorealistic molecular graphics. *Acta Cryst D* 50, 869-873.
- Miura, T., Klaus, W., Ross, A., Guntert, P., and Senn, H. (2002). The NMR structure of the class I human ubiquitin-conjugating enzyme 2b. *J Biomol NMR* 22, 89-92.
- Moraes, T. F., Edwards, R. A., McKenna, S., Pastushok, L., Xiao, W., Glover, J. N., and Ellison, M. J. (2001). Crystal structure of the human ubiquitin conjugating enzyme complex, hMms2-hUbc13. *Nat Struct Biol* 8, 669-673.
- Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu Rev Biochem* 70, 503-533.
- Sali, A., and Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234, 779-815.
- Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR

experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog NMR Spec* 34, 93-158.

Scott, W. R. P., Hunenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennel, J., Torda, A. E., Huber, T., Kruger, P., and van Gunsteren, W. F. (1999). THE GROMOS biomolecular simulation program package. *J Phys Chem* 103, 3596-3607.

Tironi, I. G., Sperb, R., Smith, P. E., and van Gunsteren, W. F. (1995). A generalized reaction field method for molecular dynamics simulations. *J Chem Phys* 102, 5451-5459.

Tong, H., Hateboer, G., Perrakis, A., Bernards, R., and Sixma, T. K. (1997). Crystal structure of murine/human Ubc9 provides insight into the variability of the ubiquitin-conjugating system. *J Biol Chem* 272, 21381-21387.

van Nuland, N. A., Hangyi, I. W., van Schaik, R. C., Berendsen, H. J., van Gunsteren, W. F., Scheek, R. M., and Robillard, G. T. (1994). The high-resolution structure of the histidine-containing phosphocarrier protein HPr from *Escherichia coli* determined by restrained molecular dynamics from nuclear magnetic resonance Overhauser effect data. *J Mol Biol* 237, 544-559.

VanDemark, A. P., Hofmann, R. M., Tsui, C., Pickart, C. M., and Wolberger, C. (2001). Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. *Cell* 105, 711-720.

Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* 8, 127-134.

Wang, G., Louis, J. M., Sondej, M., Seok, Y. J., Peterkofsky, A., and Clore, G. M. (2000). Solution structure of the phosphoryl transfer complex between the signal transducing proteins HPr and IIA(glucose) of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system. *Embo J* 19, 5635-5649.

Weissman, A. M. (2001). Themes and variations on ubiquitylation. *Nat Rev Mol Cell Biol* 2, 169-178.

Winkler, G. S., Albert, T. K., Dominguez, C., Legtenberg, Y. I., Boelens, R., and Timmers, H. T. M. (2004). An altered-specificity ubiquitin-conjugating enzyme/ubiquitin protein ligase pair. *J Mol Biol* 337, 157-165.

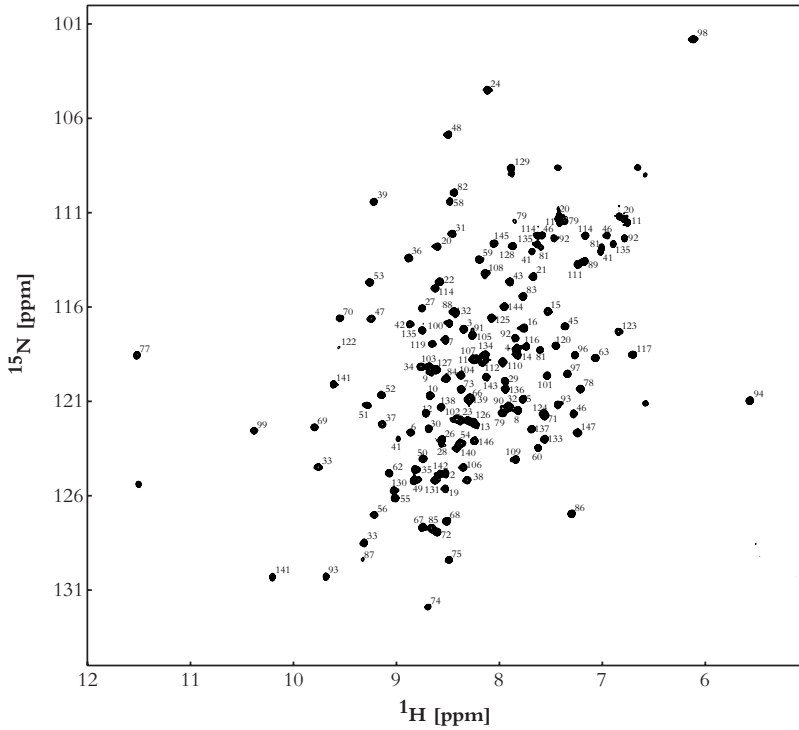
Worthylake, D., Meadow, N. D., Roseman, S., Liao, D. I., Herzberg, O., and Remington, S. J. (1991). Three-dimensional structure of the *Escherichia coli* phosphocarrier protein III_glc. *Proc Natl Acad Sci U S A* 88, 10382-10386.

Worthylake, D. K., Prakash, S., Prakash, L., and Hill, C. P. (1998). Crystal structure of the *Saccharomyces cerevisiae* ubiquitin-conjugating enzyme Rad6 at 2.6 Å resolution. *J Biol Chem* 273, 6271-6276.

Zheng, N., Wang, P., Jeffrey, P. D., and Pavletich, N. P. (2000). Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* 102, 533-539.

Appendix

(^{15}N - ^1H)-HSQC spectrum of UbcH5B at 300K, pH 7.0. Backbone assignment is indicated.



List of chemical shifts (in ppm) for UbcH5B at 300K, pH 7.0

Res	Num	N (HN)	C	C α (H α)	C β (H β)	Others
Met	1		178.7	58.5 (4.34)	32.4 (2.35, 2.35)	C γ 32.4 (2.74, 2.84)
Ala	2	124.7 (8.53)	179.6	56.0 (4.33)	19.6 (1.90)	
Leu	3	117.1 (8.36)	178.6	58.0 (3.78)	42.4 (1.80, 1.93)	C γ 32.8 (1.95), C δ 29.1, 24.1 (1.03, 0.94)
Lys	4	118.2 (7.85)	179.7	59.9 (4.14)	32.5 (2.03, 2.03)	C γ 25.2 (1.54, 1.54), C δ 29.1 (1.81, 1.81), C ϵ 36.4 (3.09, 3.09)
Ar	5	120.8 (7.79)	177.5	57.7 (4.38)	28.8 (2.11, 2.27)	H γ (1.70, 1.70)
Ile	6	122.6 (8.89)	178.2	66.9 (3.27)	38.3 (1.31)	C γ 29.8, 18.2 (1.43, 1.43, 0.58), H δ (-0.37)

Appendix

His	7	117.6 (8.54)	178.4	60.7 (4.21)	30.1 (3.14, 3.29)	
Lys	8	121.3 (7.84)	178.8	59.8 (4.21)	32.4 (2.29, 2.29)	C γ 25.2 (2.14, 2.14), C δ 25.2 (1.64, 1.64), C ϵ 42.5 (3.10, 3.10)
Ile	9	119.3 (8.68)	180.0	60.6 (4.27)	29.4 (2.38, 2.38)	
Leu	10	120.5 (8.69)	180.3	58.3 (4.25)	41.7 (1.84, 2.08)	C γ 26.9 (1.65), C δ 25.9, 23.7 (0.67, 0.59)
Asn	11	118.7 (8.18)	178.2	56.5 (4.32)	38.7 (2.35, 3.79)	N δ 111.5 (7.42, 6.76)
Asp	12	121.5 (8.73)	178.7	57.6 (4.50)	40.4 (2.89, 3.01)	
Leu	13	122.0 (8.28)	178.1	58.0 (4.06)	42.4 (1.81, 2.04)	C γ 27.3 (1.82), C δ 26.1, 25.3 (1.14, 1.04)
Ala	14	118.4 (7.84)	179.7	53.8 (4.18)	18.3 (1.57)	
Arg	15	116.1 (7.55)	176.4	57.7 (4.24)	31.2 (1.96, 1.97)	C γ 27.5 (1.76, 1.81), C δ 43.4 (3.31, 3.31)
Asp	16	117.0 (7.77)		51.8 (5.09)	41.4 (2.44, 2.78)	
Pro	17					
Pro	18		176.4	62.0	31.3	
Ala	19	125.5 (8.53)	178.2	53.9 (4.21)	18.6 (1.49)	
Gln	20	112.7 (8.62)	174.9	57.5 (4.11)	28.3 (2.43, 2.60)	C γ 35.2 (2.45, 2.35), N ϵ 111.2 (7.42, 6.83)
Cys	21	114.3 (7.69)	173.2	57.2 (5.49)	31.8 (2.70, 2.86)	
Ser	22	114.5 (8.60)	172.5	57.3 (4.67)	65.6 (3.76, 3.76)	
Ala	23	121.9 (8.40)	174.7	52.3 (5.37)	23.2 (1.57)	
Gly	24	104.4 (8.14)		45.1 (2.59, 3.75)		
Pro	25		176.6	62.6	32.5	
Val	26	122.9 (8.57)	176.7	62.6 (3.95)	31.7 (1.84)	C γ 21.1, 21.1 (0.64, 0.84)
Gly	27	116.1 (8.76)	174.0	46.5 (3.83, 3.83)		
Asp	28	123.2 (8.57)	175.2	54.2 (4.59)	40.9 (2.75, 2.83)	
Asp	29	119.9 (7.96)	176.3	53.0 (4.91)	41.4 (2.76, 3.50)	
Met	30	122.3 (8.69)	175.7	55.7 (4.40)	31.5 (1.66, 1.88)	H γ (2.55, 2.63)
Phe	31	112.0 (8.47)	175.3	59.5 (4.42)	38.4 (3.19, 3.58)	(H δ 7.39, H ϵ 7.12, H ζ 7.42)
His	32	121.2 (7.93)	172.4	54.3 (5.57)	31.7 (3.05, 3.40)	
Trp	33	124.3 (9.78)	175.8	54.2 (5.63)	33.6 (3.01, 3.16)	H δ (6.87), N ϵ 128.4 (9.33), H ζ (7.24), H η (6.87)
Gln	34	119.1 (8.79)	174.2	54.0 (5.26)	32.4 (2.05, 2.10)	C γ 34.1 (2.37, 2.44), N ϵ 111.3 (7.43, 6.80)
Ala	35	124.5 (8.84)	175.5	50.5 (5.54)	24.0 (1.40)	
Thr	36	113.3 (8.91)	173.8	60.4 (5.12)	71.5 (3.68)	H γ (0.61)
Ile	37	122.1 (9.15)	175.5	60.1 (4.73)	42.4 (1.65)	C γ 27.2, 17.5 (1.05, 1.44, 0.85), C δ 14.5 (0.73)
Met	38	125.1 (8.33)	177.0	54.2 (5.10)	32.8 (2.02, 2.21)	C γ 32.4 (2.61, 2.70)
Gly	39	110.4 (9.24)		44.3 (3.91, 4.05)		
Pro	40		177.9	63.4	32.8	
Asn	41	122.8 (9.00)	175.4	55.2 (4.46)	38.5 (2.81, 2.81)	N δ 113.1 (7.68, 7.01)
Asp	42	116.8 (8.89)	174.1	55.1 (4.60)	39.0 (2.86, 2.95)	
Ser	43	114.6 (7.91)		57.1 (5.25)	66.2 (3.98, 4.49)	
Pro	44		174.7	63.8	32.2	
Tyr	45	116.9 (7.38)	175.4	55.7 (5.38)	38.4 (2.47, 3.83)	H δ (7.28), H ϵ (6.70)
Gln	46	121.6 (7.30)	176.3	58.0 (3.77)	28.9 (2.18, 2.18)	C γ 33.4 (2.46, 2.56), N ϵ 112.2 (7.59, 6.95)
Gly	47	116.5 (9.26)	174.1	45.2 (3.70, 4.39)		
Gly	48	106.8 (8.52)	172.8	44.9 (3.35, 4.32)		
Val	49	125.1 (8.84)	173.3	62.6 (4.38)	33.3 (1.82)	C γ 21.4, 21.4 (0.57, 0.91)
Phe	50	123.9 (8.76)	174.1	56.5 (4.74)	41.5 (2.90, 3.01)	H δ (7.18), H ϵ (7.64), H ζ (7.54)
Phe	51	121.1 (9.30)	176.0	57.0 (5.35)	40.7 (3.09, 3.09)	H δ (7.24), H ϵ (7.27), H ζ (7.24)
Leu	52	120.6 (9.16)	176.8	54.6 (5.38)	46.0 (1.13, 1.40)	C γ 26.6 (1.04), C δ 25.1, 25.1 (0.28, 0.01)
Thr	53	114.6 (9.27)	173.2	60.5 (4.97)	71.4 (4.13)	H γ (1.31)

Ile	54	123.1 (8.39)	173.9	60.0 (4.67)	41.7 (1.32)	C γ 17.4 (0.66), C δ 14.6 (-0.29)
His	55	126.0 (9.03)	175.6	53.9 (5.49)	32.6 (2.89, 3.09)	H δ (6.93), H ϵ (8.31)
Phe	56	126.9 (9.23)		56.4 (4.30)	40.2 (2.44, 2.81)	H δ (7.39), H ϵ (7.31)
Pro	57		176.9	62.1	32.1	
Thr	58	110.4 (8.49)	173.6	64.8 (3.98)	68.7	H γ (1.1)
Asp	59	113.4 (8.21)	176.3	52.1 (4.98)	39.0 (2.61, 3.00)	
Tyr	60	123.4 (7.64)		58.2 (4.64)	39.9 (2.82, 3.50)	
Pro	61		175.0	63.6	32.8	
Phe	62	124.7 (9.09)	175.4	60.8 (4.40)	38.0 (3.22, 3.42)	H δ (7.40), H ϵ (7.71), H ζ (7.60)
Lys	63	118.6 (7.09)		52.1 (4.47)	34.9 (1.36, 1.79)	C γ 25.6 (1.37, 1.65), C δ 28.7 (1.83, 183), C ϵ 42.4 (3.07, 3.07)
Pro	64					
Pro	65		175.0	61.4	30.9	
Lys	66	120.7 (8.31)	175.8	55.2 (4.50)	33.1 (1.78, 1.83)	C γ 24.7 (1.34, 1.45), C δ 29.1 (1.66, 1.66), C ϵ 42.1 (2.96, 2.96)
Val	67	127.5 (8.75)	173.3	60.7 (4.89)	33.9 (1.75)	H γ (0.58, 0.97)
Ala	68	127.3 (8.53)	177.0	50.9 (5.01)	22.3 (1.29)	
Phe	69	122.4 (9.82)	177.3	59.2 (4.80)	40.2 (3.01, 3.51)	H δ (7.30), H ϵ (6.77), H ζ (6.54)
Thr	70	116.5 (9.56)	175.4	63.2 (4.50)	68.6 (4.33)	H γ (1.27)
Thr	71	121.7 (7.58)	173.8	63.7 (4.51)	71.4 (4.24)	H γ (1.17)
Arg	72	127.8 (8.62)	175.0	57.2 (4.19)	30.5 (1.24, 1.34)	C γ 26.9 (1.40, 1.56), C δ 42.8 (2.35, 2.52)
Ile	73	120.3 (8.39)	171.0	60.4 (4.80)	41.1 (1.66)	H γ (0.45, 0.55), H δ (-0.05, 0.24)
Tyr	74	131.8 (8.72)	173.3	57.9 (3.95)	38.0 (2.77, 2.77)	H δ (7.45), H ϵ (6.67)
His	75	129.2 (8.50)		55.7 (4.90)	35.6 (2.89, 2.89)	
Pro	76		177.3	64.7	32.3	
Asn	77	118.3 (11.53)		54.1 (4.20)	41.6 (1.57, 1.57)	
Ile	78	120.3 (7.24)	174.5	60.2 (5.27)	42.2 (1.10)	C γ 27.2, 22.9 (0.55, 0.82, -0.41), H δ (-0.70)
Asn	79	121.5 (7.99)		50.6 (4.96)	40.1 (2.67, 3.66)	N δ 111.5 (7.85, 7.42)
Ser	80		175.2	60.8	63.1	
Asn	81	118.1 (7.62)	175.4	53.2 (4.98)	39.0 (2.82, 3.04)	N δ 112.9 (7.60, 7.00)
Gly	82	109.8 (8.46)	174.6	46.1 (3.81, 4.99)		
Ser	83	115.3 (7.79)	172.1	59.2 (4.59)	63.5 (3.71, 3.85)	
Ile	84	119.7 (8.52)	175.8	59.6 (4.63)	41.9 (1.65)	C γ 27.2, 17.5 (0.70, 1.57, 0.81), C δ 15.3 (0.77)
Cys	85	127.6 (8.67)	171.9	57.6 (4.56)	26.2 (2.68, 3.01)	
Leu	86	126.8 (7.32)	176.9	53.9 (4.72)	46.1 (1.40, 1.61)	C γ 27.5 (1.61), C δ 24.7, 25.9 (0.98, 0.88)
Asp	87	129.4 (9.35)	180.3	58.7 (4.46)	39.4 (2.69, 2.69)	
Ile	88	116.1 (8.46)	173.4	64.9 (3.85)	37.9 (2.16)	C γ 26.8 (1.21, 1.99, 0.97), H δ (1.05)
Leu	89	113.5 (7.19)	176.8	54.1 (4.71)	41.8 (1.92, 1.96)	C γ 27.1 (1.61), C δ 25.0, 23.7 (0.91, 1.05)
Arg	90	121.2 (7.93)	176.5	56.4 (4.65)	31.4 (2.15, 2.15)	H γ (1.76, 1.72), C δ 43.8 (3.29, 3.29)
Ser	91	117.3 (8.29)	175.3	61.1 (4.52)	63.6 (4.11, 4.11)	
Gln	92	117.6 (7.86)	174.6	54.5 (4.64)	28.4 (1.35, 1.33)	C γ 33.7 (2.32, 2.42), N ϵ 112.4 (7.45, 6.81)
Trp	93	121.1 (7.45)	176.4	69.7 (4.51)	30.0 (2.91, 3.56)	H δ (6.71), N ϵ 130.3 (9.70, 7.12), H ζ (7.11, 6.81), H η (6.64)
Ser	94	120.8 (5.58)		54.6 (4.54)	65.2 (3.39, 3.92)	
Pro	95		175.4	63.7	31.6	
Ala	96	118.4 (7.29)	178.9	52.6 (4.13)	18.8 (1.21)	
Leu	97	119.4 (7.36)	173.1	54.8 (4.04)	42.4 (1.10, 1.10)	C γ 27.2 (1.38), C δ 27.2, 27.2 (0.76, 0.85)
Thr	98	101.7 (6.13)	176.4	58.1 (5.17)	73.1	H γ (1.22)
Ile	99	122.5 (10.41)		61.0 (4.12)	36.2 (2.17)	C γ 27.8, 19.0 (1.22, 1.56, 1.08), H δ (0.49)
Ser	100	116.8 (8.50)	175.6	62.6 (3.97)	(4.04, 4.04)	

Appendix

Lys	101	119.5 (7.55)	180.2	58.9 (4.14)	32.3 (2.04, 2.04)	Cγ 25.6 (1.58, 1.58), Cδ 29.0 (1.74, 1.74), Cε 42.0 (3.00, 3.00)
Val	102	121.8 (8.43)	177.5	66.9 (3.59)	31.6 (2.51)	Cγ 21.3, 23.1 (0.85, 1.27)
Leu	103	119.0 (8.70)	179.3	58.6 (3.93)	40.9 (1.91, 1.91)	Hγ (2.04), Hδ (0.80, 0.64)
Leu	104	119.5 (8.39)	180.4	58.6 (4.12)	41.7 (1.67, 1.94)	Cγ 27.1 (1.88), Cδ 23.6, 23.6 (1.00, 0.97)
Ser	105	117.4 (8.28)	177.3	62.6 (4.37)	63.8 (4.00, 4.00)	
Ile	106	124.4 (8.37)	177.2	66.1 (3.58)	37.5 (2.02)	Cγ 30.3, 18.0 (2.03, 2.03, 0.82), Cδ 14.9 (0.64)
Cys	107	118.7 (8.28)	177.3	65.4 (3.93)	26.2 (2.94, 3.29)	
Ser	108	114.2 (8.16)	176.2	62.0 (4.31)	62.7 (4.10, 4.10)	
Leu	109	124.0 (7.86)	179.9	56.6 (4.19)	42.5 (1.65, 2.04)	Cγ 26.8 (1.55), Cδ 25.7, 23.4 (0.72, 1.06)
Leu	110	118.8 (7.98)	177.4	58.8 (3.94)	41.1 (1.58, 2.10)	Hγ (1.41), Cδ 26.2, 21.7 (0.79, 0.64)
Cys	111	113.6 (7.26)	178.8	61.7 (4.41)	27.9 (3.17, 3.17)	
Asp	112	118.8 (8.19)		58.3 (5.17)	38.5 (2.49, 2.68)	
Pro	113		173.9	62.4	32.6	
Asn	114	114.9 (8.64)		49.2 (5.39)	38.9 (2.81, 2.99)	Nδ 112.2 (7.63, 7.16)
Pro	115		174.3	63.7	32.2	
Asp	116	118.0 (7.75)	175.3	55.4 (4.63)	40.6 (2.65, 2.81)	
Asp	117	118.4 (6.72)		51.5 (4.96)	41.1 (2.24, 2.76)	
Pro	118		177.1	62.6	33.1	
Leu	119	117.9 (8.67)	177.2	55.4 (4.74)	44.2 (1.41, 1.60)	Cγ 27.0 (1.62), Cδ 25.0, 25.0 (0.87, 0.87)
Val	120	118.0 (7.47)		59.0 (4.72)	31.9 (2.22)	Cγ 19.9, 23.1 (0.96, 1.10)
Pro	121		178.6	65.4	32.6	
Glu	122	118.1 (9.58)	178.5	59.8 (4.21)	29.3 (1.98, 2.20)	Cγ 36.5 (2.35, 2.42)
Ile	123	117.2 (6.86)	178.0	65.2 (3.74)	38.7 (1.46)	Cγ 21.0, 17.4 (1.31, 1.31, 0.35), Cδ 14.29 (0.22)
Ala	124	121.6 (7.59)	177.9	55.6 (3.72)	18.7 (1.60)	
Arg	125	116.4 (8.10)	179.0	59.8 (4.09)	30.1 (1.96, 1.96)	Cγ 27.5 (1.60, 1.77), Cδ 43.3 (3.29, 3.29)
Ile	126	121.9 (8.32)	176.7	65.3 (3.54)	38.0 (0.99)	Cγ 28.4, 16.6 (1.66, 1.66, 0.80), Cδ 13.8 (-0.36)
Tyr	127	119.2 (8.62)	176.7	60.5 (3.38)	38.4 (2.46, 2.93)	
Lys	128	112.7 (7.89)	178.5	59.2 (4.01)	33.7 (1.95, 1.95)	Cγ 25.9 (1.55, 1.55), Cδ 29.3 (1.68, 1.68), Cε 41.5 (2.93, 2.93)
Thr	129	108.6 (7.90)	174.9	63.1 (4.44)	71.5 (4.21)	Hγ (1.32)
Asp	130	125.6 (9.04)	174.1	52.9 (4.78)	40.5 (2.49, 2.96)	
Arg	131	125.1 (8.64)	178.1	57.4 (4.77)	28.7 (1.81, 1.81)	Cγ 25.6 (1.49, 1.49), Cδ 42.1 (3.00, 3.00)
Glu	132	116.3 (8.44)	179.2	59.4 (4.18)	29.0 (2.12, 2.12)	Cγ 36.4 (2.33, 2.43)
Lys	133	122.9 (7.58)	177.5	59.7 (3.96)	32.4 (1.87, 1.87)	
Tyr	134	118.5 (8.17)	175.7	62.2 (4.50)	38.4 (2.95, 3.02)	
Asn	135	117.1 (8.77)	177.6	55.8 (4.09)	37.8 (2.75, 2.98)	Nδ 112.7 (7.64, 6.89)
Arg	136	120.2 (7.96)	179.9	59.9 (3.99)	30.2 (2.00, 2.00)	Cγ 28.8 (1.59, 1.80), Cδ 43.7 (2.96, 2.96)
Ile	137	122.4 (7.70)	177.6	65.3 (3.73)	38.0 (2.17)	Cγ 29.4, 17.6 (1.32, 1.32, 1.17), Cδ 12.7 (0.96)
Ala	138	121.2 (8.58)		55.8 (3.95)	17.4 (1.11)	
Arg	139	120.8 (8.32)	178.8	60.1 (3.99)	30.4 (2.04, 2.04)	
Glu	140	123.4 (8.43)	180.4	59.9 (4.14)	29.4 (2.33, 2.46)	Cγ 36.6 (2.33, 2.59)
Trp	141	120.0 (9.63)	181.0	61.6 (4.44)	29.0 (3.61, 3.76)	Hδ (7.32), Nε 130.3 (10.22, 7.35), Hζ (7.71, 7.09), Hη (7.43)
Thr	142	124.8 (8.60)	175.8	68.6 (4.38)		Hγ (1.34)
Gln	143	119.6 (8.16)	177.5	58.5 (4.23)	28.5 (2.14, 2.26)	Cγ 34.0 (2.49, 2.59)
Lys	144	115.8 (7.97)	178.5	58.8 (3.93)	33.4 (0.91, 1.23)	Cγ 25.0 (0.66, 1.07), Cδ 29.4 (1.18, 1.27), Cε 41.9 (2.72, 2.72)
Tyr	145	112.6 (8.07)	176.3	58.0 (5.08)	40.6 (2.80, 3.61)	(Hδ 7.22, Hε 6.01)
Ala	146	123.0 (8.26)	175.2	51.4 (4.95)	21.6 (1.11)	
Met	147	122.6 (7.25)		57.0 (4.58)	35.5 (1.97, 2.27)	Cγ 33.4 (2.56, 2.56)

Summary

Understanding the molecular and functional interactions among macromolecular complexes, as well as their changes associated with time, cell type or disease state will be invaluable for human health, and will have direct implications, for example, in pharmaceutical research to identify and select potential targets, and design efficient and specific drugs. Structural studies of macromolecular complexes, however, suffer from some limitations, especially in the case of weak and transient complexes. Because of that, it will not be possible to study all macromolecular interactions at an atomic level. New and complementary methodologies, such as docking, have therefore been developed. The current computational approaches, however, also suffer from limitations and new developments and improvements are needed.

This thesis, as outlined in a short general introduction, describes structural studies on the UbcH5B-CNOT4 complex involved in the ubiquitination pathway and furthermore introduces a new docking approach in which the docking of two macromolecules is driven by biophysical and/or biochemical information.

Chapter 1 gives an introduction to the RING motif and its role in the ubiquitination pathway. Structural and chemical properties as well as biological and medical implications of the RING finger domain family of proteins and a number of related systems are described. Differences in the structure of various RING domains and related domains are also discussed.

Chapter 2 focuses on structural studies of macromolecular complexes using NMR and docking approaches. First, the structure determination of complexes by NMR methods is presented, various restraints commonly used in NMR to study complexes are described and limitations are discussed. Secondly, a general overview of docking approaches is described. Finally, new methods that combine NMR restraints and docking approaches are described.

Chapter 3 describes HADDOCK, a new docking approach that allows for flexibility at the interface and that is based on biophysical and/or biochemical information. This information, derived for example from NMR chemical shift perturbation or site-directed mutagenesis experiments, is converted into highly ambiguous intermolecular distance restraints that are directly used to drive the docking process. The docking protocol allows for side chains and backbone flexibility at the interface and the solutions are scored according to an intermolecular energy term. The method is tested on three complexes. For two of these complexes, intermolecular distance information is derived from NMR chemical shift mapping and for the third complex, derived from mutagenesis data. We demonstrate that, in all cases, the lowest energy structures generated by HADDOCK are the closest from the original structure.

Chapter 4 describes the solution structure of UbcH5B, an E2 ubiquitin conjugating enzyme, which is solved using a combination of homology modeling, NMR automated NOE assignment and diffusion anisotropy data. UbcH5B is 90% homologous to Ubc4 for

which an X-ray structure is available. We therefore created a homology model of UbcH5B based on the Ubc4 structure. NMR relaxation measurements are performed on UbcH5B. They show limited motions for the major part of the protein backbone. This information is translated into diffusion anisotropy restraints and used to refine the homology model. This refined model serves as starting point to the automated NOE assignment process. The final structure is well defined. We compare the structure of UbcH5B with other E2 structures. The global fold of all E2s is very similar. Some differences are, however, observed and correlate well with the dynamical properties of E2s. Importantly, the position and orientation of the N-terminal α -helix as compared to the core of the protein differ in the various structures. This is in agreement with our observation that the loop that connects this helix to the core of the protein is flexible. This flexibility may be determinant in E3 ubiquitin ligase binding and recognition. Furthermore, a highly conserved asparagine residue was previously shown to be important for the ubiquitin transfer by catalyzing an isopeptide bond formation. In crystal structures, this asparagine, however, points away from the active site cysteine and can not accomplish its catalytic role. Analysis of the structure of UbcH5B and another Ubc structure also solved by NMR shows that in solution, this asparagine is in close proximity to the active site cysteine, in a conformation suitable for its catalytic role.

In Chapter 5, HADDOCK is used to generate a structural model of the UbcH5B-CNOT4 complex. It has previously been reported that CNOT4 is involved in the ubiquitination pathway as an E3 ubiquitin ligase and specifically interacts with UbcH5B, an E2 ubiquitin conjugating enzyme. The residues of the CNOT4 RING domain important for the interaction with UbcH5B were previously reported. In this Chapter, the residues of UbcH5B important for the binding to CNOT4 RING are identified from NMR chemical shift perturbation experiments. These data are used to generate a structural model of the UbcH5B-CNOT4 complex. Two sets of solutions are, however, obtained that can not be discriminated. Based on these docking results, mutagenesis experiments are performed and identify several basic residues of UbcH5B important for the binding to CNOT4. More importantly, charge-altered mutants of CNOT4 (Glu49Lys) and of UbcH5B (Lys63Glu) can restore the binding indicating an electrostatic interaction between these two residues. Once this information is included in the docking, a unique set of solutions is obtained. The structural model of the UbcH5B-CNOT4 complex is compared with the X-ray structure of the homologous UbcH7-c-Cbl complex and significant differences at specific residues give structural insights into the mechanisms of the E2-E3 specificity.

Samenvatting

Inzicht in de interacties en functies van macromoleculaire complexen en in hun veranderingen in de tijd en in relatie tot celtype of ziektestadium is belangrijk voor medisch onderzoek. Dit kan zeer nuttige informatie geven voor het selecteren van targets voor de ontwikkeling van specifieke en efficiënte medicijnen. Het structuuronderzoek aan deze macromoleculaire complexen is echter lastig, vooral als het gaat om zwakke interacties en kortlevende complexen. Er zijn daarom computermethoden voor het bepalen van structuren van onder andere eiwit-eiwit complexen ontwikkeld, maar aangezien deze dokkingmethoden ook hun beperkingen hebben, zijn verbeteringen en vernieuwingen in dit gebied nog steeds nodig.

In dit proefschrift wordt een studie beschreven naar de structuur van het UbcH5B-CNOT4 complex dat een rol speelt in ubiquitineren. Daarnaast wordt een nieuwe methode om twee macromoleculen te dokken belicht, waarbij biofysische en/of biochemische informatie de drijvende kracht is. Een korte introductie is gegeven in de algemene inleiding.

Hoofdstuk 1 geeft een introductie in het RING-vinger motief en de rol dat het speelt in het ubiquitineringsproces. Zowel de structuur en chemische eigenschappen van de eiwitten in de RING-vinger familie en vergelijkbare systemen, als ook de biologische en medische toepassingen worden beschreven. Ook worden de verschillen in structuur van een aantal RING- en vergelijkbare domeinen besproken.

Hoofdstuk 2 concentreert zich op het bepalen van structuren van macromoleculaire complexen met zowel NMR spectroscopie als dokkingmethoden. Als eerste wordt ingegaan op de structuurbevestiging van eiwitcomplexen met behulp van NMR. Verschillende experimentele NMR data, die worden gebruikt bij het bestuderen van complexen met NMR, en hun beperkingen worden besproken. Ten tweede wordt een algemeen overzicht gegeven van bestaande dokkingmethoden. Als laatste worden nieuwe methoden beschreven die NMR informatie en dokken combineren.

HADDOCK, een nieuwe dokkingmethode die flexibiliteit aan het complexgrensvlak toestaat en gebaseerd is op biofysische en/of biochemische informatie, wordt in hoofdstuk 3 beschreven. Deze informatie, verkregen uit bijvoorbeeld veranderingen in NMR chemische verschuivingen of specifieke mutatie experimenten, wordt vertaald in zeer ambigu intermoleculaire afstanden, die direct gebruikt worden om het dokkingsproces te leiden. Het dokkingprotocol staat flexibiliteit van zowel de hoofdketen als de zijketens aan het complexgrensvlak toe en de oplossingen worden gescoord op basis van intermoleculaire energie. De methode is getest op drie complexen. In het geval van twee van deze complexen was de intermoleculaire afstands-informatie gebaseerd op veranderingen in NMR chemische verschuivingen en voor het derde complex op specifieke mutatie data. We laten zien dat in alle gevallen de HADDOCK structuren met de laagste energie het beste overeenstemmen met de werkelijke structuur.

In hoofdstuk 4 wordt de NMR structuur van UbcH5B, een E2 ubiquitine conjugerend enzym, besproken. Voor de structuurbevestiging is gebruik gemaakt van een homoloog

model, geautomatiseerd toekennen van NOEs en diffusie-anisotropie data. UbcH5B is 90% homologo aan Ubc4, waarvan een kristalstructuur bekend is. Op basis van de Ubc4 structuur werd daarom een homologo model voor UbcH5B gecreëerd. Er werden NMR relaxatie-experimenten uitgevoerd met UbcH5B. Deze laten zien dat het overgrote deel van de eiwitheofdketen weinig beweeglijk is. Deze informatie is vertaald naar diffusie-anisotropie data en is gebruikt om het homologe model te verfijnen. Dit verfijnde model werd vervolgens gebruikt als startpunt voor het geautomatiseerde NOE-toekenningsproces. De uiteindelijke structuur is goed gedefinieerd. We vergelijken de structuur van UbcH5B met andere E2 structuren en de globale vouwing van alle E2's is erg vergelijkbaar. Er zijn echter een aantal verschillen aanwezig, die correleren met de dynamische eigenschappen van E2's. Voornamelijk de positie en oriëntatie van de N-terminale α -helix ten opzichte van de kern van het eiwit zijn anders in de verschillende structuren. Dit stemt overeen met de waarneming dat de lus, die deze helix met de kern van het eiwit verbindt, flexibel is. Deze flexibiliteit zou bepalend kunnen zijn voor de interactie met E3 ubiquitine ligases. Verder is in het verleden aangetoond dat een zeer geconserveerd asparagine residu erg belangrijk is voor het doorgeven van ubiquitine, omdat het de vorming van een isopeptidebinding katalyseert. In alle E2 kristalstructuren heeft deze asparagine echter een oriëntatie weggedraaid van de cysteïne in het actieve centrum en kan in deze positie niet zijn katalyserende rol vervullen. Analyse van zowel de UbcH5B structuur als een andere Ubc structuur, die ook met NMR werd bepaald, laat zien dat in oplossing deze asparagine zich in de nabijheid van de cysteïne in het actieve centrum bevindt, in een conformatie die zich dus wel leent voor katalyse.

In hoofdstuk 5 is HADDOCK gebruikt om een structuurmodel voor het UbcH5B-CNOT4 complex te genereren. Eerder is reeds beschreven dat CNOT4 de rol van een E3 ubiquitine ligase heeft in het ubiquitineringsproces en specifiek een interactie aangaat met UbcH5B, dat een E2 ubiquitine conjugerend enzym is. De residuen in het CNOT4 RING-domein die belangrijk zijn voor de interactie met UbcH5B werden reeds elders gerapporteerd. Op basis van veranderingen in NMR chemische verschuivingen zijn in dit hoofdstuk de residuen van UbcH5B die belangrijk zijn voor de binding met het RING-domein van CNOT4 geïdentificeerd. Deze data zijn gebruikt om een structuurmodel voor het UbcH5B-CNOT4 complex te genereren. In eerste instantie werden echter twee clusters van oplossingen verkregen, waartussen geen onderscheid gemaakt kon worden. Gebaseerd op deze resultaten zijn specifieke mutatie experimenten uitgevoerd, waaruit blijkt dat een aantal basische residuen in UbcH5B belangrijk zijn voor de interactie met CNOT4. Daarnaast kunnen de mutanten van CNOT4 (Glu49Lys) en UbcH5B (Lys63Glu) waarin een omkering van lading is aangebracht de binding herstellen, hetgeen betekent dat er een elektrostatistische interactie tussen de gemuteerde residuen moet zijn. Door deze informatie toe te voegen aan het dokkingprotocol wordt één unieke set van oplossingen verkregen. Het model voor het complex tussen UbcH5B en CNOT4 wordt vergeleken met de kristalstructuur van het homologe UbcH7-c-Cbl complex. Een aantal specifieke en significante verschillen geven inzicht in het mechanisme van de E2-E3 specificiteit.

Résumé

La compréhension des interactions moléculaires et fonctionnelles parmi les complexes macromoléculaires, ainsi que leurs changements associés au temps, au type cellulaire ou à un état pathologique sera inestimable pour la recherche médicale et aura des implications directes, par exemple, dans la recherche pharmaceutique pour identifier et sélectionner des cibles potentielles, et créer des médicaments efficaces et spécifiques. Toutefois, les études structurales de complexes macromoléculaires souffrent de certaines limitations, en particulier dans le cas de complexes transitoires. A cause de cela, il ne sera pas possible d'étudier toutes les interactions macromoléculaires à un niveau atomique. Des méthodologies nouvelles et complémentaires, comme l'arrimage moléculaire, ont donc été développées. Toutefois, ces approches informatiques actuelles souffrent aussi de limitations et de nouveaux développements sont nécessaires.

Cette thèse, comme indiqué dans une courte introduction générale, décrit l'étude structurale du complexe UbcH5B-CNOT4 impliqué dans la voie de l'ubiquitination et, en outre, introduit une nouvelle approche d'arrimage moléculaire dans laquelle l'arrimage est dirigé par des informations biochimiques ou biophysiques.

Le premier chapitre présente une introduction sur le motif protéique 'RING' et son rôle dans la voie de l'ubiquitination. Les propriétés structurales et chimiques ainsi que les implications biologiques et médicales de la famille de protéines contenant le domaine 'RING' et d'un nombre de domaines apparentés sont décrites. Les différences de structure entre plusieurs domaines RING et d'autres domaines sont discutées.

Le deuxième chapitre se concentre sur les études structurales de complexes macromoléculaires utilisant la RMN et des approches d'arrimage moléculaire. D'abord, la détermination de structures de complexes par RMN est présentée, diverses contraintes communément utilisées en RMN pour étudier des complexes sont décrites et les limitations sont discutées. ensuite, une vue générale des approches d'arrimage moléculaire est présentée. enfin, de nouvelles méthodes qui combinent des contraintes RMN et des approches d'arrimage moléculaire sont décrites.

Le troisième chapitre décrit HADDOCK, une nouvelle approche d'arrimage moléculaire permettant l'introduction de flexibilité à l'interface et basée sur des informations biophysiques et/ou biochimiques. Ces informations, dérivées de perturbation de déplacements chimiques par RMN ou de données de mutagenèse dirigée, sont converties en contraintes de distances intermoléculaires ambiguës qui sont utilisées directement pour diriger le processus d'arrimage moléculaire. Le protocole d'arrimage permet une flexibilité des chaînes latérales et de la chaîne principale à l'interface et les solutions sont classées en fonction d'un terme d'énergie intermoléculaire. La méthode est testée sur trois complexes. Pour deux de ces complexes, l'information de distance intermoléculaire est dérivée de données de perturbation de déplacements chimiques par RMN et pour le troisième complexe, dérivée de données de mutagenèse dirigée. Nous démontrons que dans tous les cas, les structures de plus faible énergie générées par HADDOCK sont les plus proches de la structure originale.

Le quatrième chapitre décrit la structure en solution d'UbcH5B, une enzyme de conjugaison de l'ubiquitine (E2) résolue en utilisant une combinaison de modèle par homologie, d'attribution automatique des NOEs et de données d'anisotropie de diffusion. UbcH5B possède 90% d'homologie avec Ubc4 dont la structure cristallographique est connue. Nous avons donc créé un modèle par homologie de UbcH5B basé sur la structure d'Ubc4. Des mesures de relaxation par RMN sont effectuées sur UbcH5B. Ces mesures montrent des mouvements limités pour la majeure partie de la chaîne principale de la protéine. Cette information, traduite en contraintes d'anisotropie de diffusion est utilisée pour affiner le modèle par homologie. Ce modèle affiné sert de point de départ pour le processus d'attribution automatique des NOEs. La structure finale est bien définie. Nous comparons la structure d'UbcH5B avec celle d'autres enzymes E2 et le repliement global est très similaire. Toutefois, quelques différences sont observées et sont en accord avec les propriétés dynamiques des E2s. La position et l'orientation de l'hélice N-terminale par rapport au coeur de la protéine sont différentes dans les diverses structures. Ceci est en accord avec notre observation que la boucle qui connecte cette hélice au coeur de la protéine est flexible. Cette flexibilité peut être importante dans la reconnaissance et la fixation des ligases de l'ubiquitine E3s. De plus, il a été montré précédemment qu'une asparagine très conservée est importante pour le transfert de l'ubiquitine en catalysant la formation de la liaison isopeptidique. Dans les structures cristallographiques, cette asparagine pointe dans une direction opposée à la cystéine du site actif et ne peut pas accomplir son rôle catalytique. L'analyse de la structure d'UbcH5B et d'une autre structure d'Ubc résolue aussi par RMN montre que, en solution, cette asparagine est proche du site actif, dans une conformation favorable pour son rôle catalytique.

Dans le cinquième chapitre, HADDOCK est utilisé pour générer un modèle structural du complexe UbcH5B-CNOT4. Il a été démontré précédemment que CNOT4 est impliqué dans la voie de l'ubiquitination en tant que ligase de l'ubiquitine (E3) et interagit de manière spécifique avec UbcH5B, une enzyme de conjugaison de l'ubiquitine (E2). Les résidus du domaine 'RING' de CNOT4 importants pour l'interaction ont été définis dans une étude précédente. Dans ce chapitre, les résidus d'UbcH5B importants pour l'interaction avec le domaine 'RING' de CNOT4 sont identifiés par des expériences de perturbation des déplacements chimiques par RMN. Ces données sont utilisées pour générer un modèle structural du complexe UbcH5B-CNOT4. Toutefois, deux ensembles de solutions sont obtenus et ne peuvent être discriminés. Basées sur ces résultats d'arrimage moléculaire, des expériences de mutagenèse dirigée sont réalisées et permettent d'identifier plusieurs résidus basiques d'UbcH5B importants pour l'interaction avec CNOT4. De plus, des mutants avec une inversion de charge de CNOT4 (Glu49Lys) et d'UbcH5B (Lys63Glu) restaurent l'interaction indiquant une interaction électrostatique entre ces deux résidus. Une fois cette information introduite dans l'arrimage moléculaire, un ensemble unique de solutions est obtenu. Le modèle structural d'UbcH5B-CNOT4 est comparé à la structure cristallographique du complexe homologue UbcH7-c-Cbl et des différences significatives impliquant des résidus spécifiques donnent des explications structurales sur les mécanismes de spécificité dans les complexes E2-E3.

Acknowledgements

Introduction

If I remember, a bit more than 4 years ago when I was still in Marseille, I would never have thought of leaving Marseille, especially to go to the big North... But well, here I am, spending more than 4 years in the Netherlands. During this period, I had the chance to meet people from all over the world. It was really fantastic. Now I am quite good in geography, well, at least I know that there is something else outside Marseille. I am therefore grateful to all the people I met during this period for all the enrichment they provided. I can evidently not cite everybody here so I have to make a selection. Sorry for the ones for which the names are not on this page, but I won't forget you anyway.

Results and Discussion

The lab in Marseille:

I first would like to thank my previous lab in Marseille, especially Françoise, Xavier, Olivier, Corinne, Brigitte, Isabelle, Cathy and Alain, first, for all I learnt during one year, and, secondly, for pushing me to come here in the Netherlands.

The lab in Utrecht:

Here there are quite a lot of people that I want to thank, basically everybody that has been or is in the lab. First of all, I would like to thank Rolf and Rob for accepting this poor Marseillais that didn't speak a word of English to join the lab. I'm still amazed that after my interview you accepted me (well, I discovered later that at this time you had so much money and so few applicants that anybody was accepted. That makes things a bit more clear). I also want to thank Rolf for saving my PhD by giving me, after one year, a new subject that worked quite well, and for everything he taught me during my PhD. Then I want to thank Alexandre for becoming my co-promotor and getting a wonderful idea of making our own docking program. I would never have thought about it and it was a high level scientific idea. Alex, thanks also for all the discussions we had during the structure calculation, docking and other things. There are also three other persons of the lab that I want to acknowledge in details: Gert, thank you for all the help and the teaching you gave me in the wet lab. I really enjoyed my first year working with you. It was a great experience. Thanks also, Albert, for your kindness and your help in the lab. Klaartje, I thank you for so many things and even more. I am so happy that we have been involved in the same project. It was really great to work together (even up to 2 in the morning). I really hope that we will keep contact for a long long time (well, Grenoble is not far from Zürich...). Now I want to acknowledge all the rest of the lab, in particular the old fellows that shared so many things with me (I still miss our PhD dinners). Thanks to: Leo, for your patience when we were speaking English at the very beginning using the French-English dictionary every 2 minutes. Antonio, you didn't stay long but it was a great time. Henry, my roommate, for the introduction to Hertog Jan. Nocky, for all these nice evenings and a nice canoe trip in Canada. Danny, especially for this excellent trip to China, and for the chili fondue (I will never forget this one and I guess Kathrin and Sacha won't either). Roberto, the banana

guru, I hope to visit Brazil once. Eugene, drinking vodka with you is always a pleasure (if I forget the headache the day after). Simona, the pasta specialist. Patrick, Ju, Ced, Audrey, and Adrien (see the french mafia section). And then there are all the others that shared many good moments with me (scientifically or around a beer): Snjezana, Monique, Bernd, Eiso, Tammo, Aalt-Jan, Aart, Michiel, Johan, Nico, Rainer, Hans, Kostas, and all the others. I also want to thank my scientific partners. Marc, I really enjoyed our collaboration. It worked very well and I really liked all our discussions. Richard, thanks for all the time you spend in the lab for me. I know, in NMR, we are asking a lot but you always found some time. Thanks also, Bas, for your input in the project.

The french mafia:

Here, I would like to thank all the french that I met in Utrecht. It was great during these four years to speak english and meet all kind of foreigners, but speaking french around some french food and some good wine is a must. Un très grand merci à Flo, la première française que j'ai rencontrée, et du sud en plus. Vincent, j'oublierai pas la tarte au maroilles et la Trois Monts. Patrick, vive la et surtout le Champagne. Julien, pour toute ces cuites qu'on s'est prises. Ced, je suis vraiment content que tu sois venu ici, on s'est bien éclaté. Adrien, Laure et Gabriel, on se fait une fondue de fromage quand vous voulez. Carine, on se fait un petit restau à Paris quand tu veux. Claire, garde ta bonne humeur, ne change pas. Je veux aussi remercier Audrey, Philippe, Steve, Alex, Laurence et Stephanie pour toutes ces bonnes soirées bien françaises.

The Family:

Je tiens a remercier toute ma famille ainsi que celle de Sandrine pour les encouragements durant ces dernières années et pour l'accueil toujours très chaleureux quand je redescends. Merci surtout à mes parents, à Emmanuelle, et à mes grand-parents. Je vous promets que dès qu'on peut, on redescend sur Marseille (on peut toujours rêver d'avoir un jour un poste à Marseille).

The Marseillais Friends:

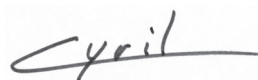
Là, ca va être rapide, il y en a peu, mais ceux là sont très importants. D'abord, gros gros gros bisous a Mag, Fab et Elora. Je nous interdit de perdre un jour contact. Bisous aussi à Isa et Soizic. Il faut qu'on se fasse une bonne bouffe quand je redescends.

The most important:

Last but not least, My love, My muse, My Sandrine. On a galéré deux ans, toi à Marseille, moi ici, mais on a tenu bon. Notre amour est plus fort que ça, même si personne n'y croyait (à part nous deux bien sûr). Merci pour tout, je ne vais pas rentrer dans le détail évidemment. La seule chose que j'ai à dire, c'est que je t'aime, I love you, Ik hou van jou...

Conclusions

These four years and a bit more where the best experience I ever got in my life. I am so glad I had the courage to come here. And all this good time I had is because of you all. So thanks again for everything, and I definitely want to keep contact with all of you. You have to promise that we will meet again, wherever in the world, but we'll meet again.



Curriculum Vitae

Cyril Dominguez was born the 6th of August 1975, in Marseille, France. He got his high school diploma in 1994, from Lycée Saint-Exupéry, Marseille. In the same year, he entered the University of Provence (Aix-Marseille I) to study biochemistry. In 1996, he joined the University of the Méditerranée (Aix-Marseille II) to study biochemistry and biophysics. In 1998, he studied nutritional biochemistry at the University of Aix-Marseille III, and carried out a one-year project in the laboratory of Dr. Françoise Guerlesquin, BIP, CNRS, Marseille. He received his pre-doctoral degree in biochemistry in September 1999 from the University of Aix-Marseille III. In January 2000, he joined the laboratory of Prof. Dr. Rob Kaptein and Prof. Dr. Rolf Boelens, to work on the PhD project described in this thesis. He will present his PhD thesis at the University of Utrecht the 14th of June, 2004. From the 1st of August 2004, Cyril Dominguez will start a post-doctoral fellow with Prof. Dr. Frédéric Allain, at the Swiss Federal Institute of Technology (ETH), in Zürich, Switzerland.

List of Publications

- 6 – C. Dominguez,* K. Houben,* F. M. A. van Schaik, H. Th. M. Timmers, A. M. J. J. Bonvin, and R. Boelens (2004). NMR solution structure of the E2 ubiquitin conjugating enzyme UbcH5B. To be submitted
- 5 – C. Dominguez, A. M. J. J. Bonvin, G. S. Winkler, F. M. A. van Schaik, H. Th. M. Timmers, and R. Boelens (2004). Structural model of the UbcH5B/CNOT4 complex revealed by combining NMR, mutagenesis and docking approaches. *Structure*, 12(4), 633-644
- 4 – G. S. Winkler, T. K. Albert, C. Dominguez, Y. I. A. Legtenberg, R. Boelens, and H. Th. M. Timmers (2004). An altered-specificity ubiquitin-conjugating enzyme/ubiquitin protein ligase pair. *J. Mol. Biol.*, 337(1), 157-165
- 3 – C. Dominguez, G. E. Folkers, and R. Boelens (2003). RING domain proteins. Contribution to Handbook of metalloproteins, Volume 3, John Wiley and Sons, 338-351
- 2 – C. Dominguez, R. Boelens, and A. M. J. J. Bonvin (2003). HADDOCK: a protein-protein docking approach based on biochemical and/or biophysical information. *J. Am. Chem. Soc.*, 125(7), pp 1731-1737
- 1 – C. Dominguez, C. Sebban-Kreuzer, O. Bornet, B. Kerfelec, C. Chapus, and F. Guerlesquin (2000). Interaction of bile salt micelles and colipase studied through intermolecular nOes. *FEBS letters*, 482(1-2), pp 109-112

* These authors contributed equally to this work

