

**NMR Studies of the Yeast Splicing Factor Prp40:
Structures and Ligand Recognition of a
WW Domain Pair and an FF Domain.**

Inaugural-Dissertation

zur Erlangung der Doktorwürde

eingereicht am Institut für Chemie
des Fachbereichs Biologie, Chemie und Pharmazie
der Freien Universität Berlin

vorgelegt von
Silke Wiesner
aus Berlin

September 2002

Eingereicht am: 24.09.2002

Datum der Disputation: 25.02.2003

1. Gutachter: Prof. Dr. H. Oschkinat and
 Institut für Chemie Forschungsinstitut für Molekulare Pharmakologie
 Freie Universität Berlin Campus Berlin-Buch
 Takustr. 3 Robert-Rössle-Str. 10
 14195 Berlin, Germany 13125 Berlin, Germany

2. Gutachter: Dr. M. J. Macias
 Institut de Recerca Biomedica
 Parc Cientific de Barcelona
 Josep Samitier 1-5
 08028 Barcelona, Spain

Weitere Kommissionsmitglieder: Prof. Dr. W. Saenger, Prof. Dr. F. Hucho

This work has been carried out from September 1998 – June 2002 under the supervision of Dr. Maria J. Macias and Dr. Michael Sattler at the EMBL Heidelberg, Germany. Parts of this Thesis have been published or will be published in the near future:

Macias M.J., Wiesner S. & Sudol M. (2002). WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett., Special Issue* **513**: 30 – 37

Wiesner S., Stier G., Sattler M. & Macias, M.J. (2002) Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. *J. Mol. Biol.* **324**: 807 – 822

Wiesner S.*, Gasch A.* & Macias M.J. Solution structure of the N-terminal FF domain of the yeast splicing factor Prp40. *Manuscript*

Publications, which are not part of this Thesis:

Wiesner S., Kurian E., Prendergast F.G. & Halle B. (1999). Water molecules in the binding cavity of intestinal fatty acid binding protein: dynamic characterization by water ^{17}O and ^2H magnetic relaxation dispersion. *J. Mol. Biol.* **286**: 233 – 246.

*Both authors contributed equally to this work.

*Wenn man das zierlichste Näschen
Von seiner liebsten Braut
Durch ein Vergrößerungsgläschen
Näher beschaut,
Dann zeigen sich haarige Berge,
Daß es einem graut.*

Genau besehn, Joachim Ringelnatz

Summary

Many eukaryotic proteins encompass multiple cooperating domains and/or domain repeats fine-tuning the functions of these so-called mosaic proteins. It is therefore of key interest to know not only the structures of the individual domains constituting a protein, but also their relative arrangement. The splicing factor Prp40 (pre-mRNA processing protein), which has been studied in this Thesis, comprises an N-terminal WW domain pair and six consecutive FF domains. In the yeast spliceosome, Prp40 is associated with the U1 snRNP (small nuclear ribonucleoprotein) and interacts with a number of splicing factors functioning in cross-intron bridging. While the Prp40 WW domains interact with proline-rich regions present in the branch-point binding protein (BBP) and the U5 snRNP component Prp8, the FF domains interact with tetratricopeptide repeats (TPR) present in the crooked neck-like factor 1 (Clf1). Furthermore, binding of Prp40 to the phosphorylated C-terminal domain (CTD) of the largest subunit of RNA polymerase II has suggested a function of Prp40 in coupling splicing to transcription.

In this Thesis, first an attempt has been made to classify members of the WW domain family based on their sequences, three-dimensional structures and proline-rich binding motifs. This led to a subdivision of WW domains in four groups. WW domains in group I recognise PPxY motifs (where x is any residue), while group II WW domains are rich in aromatic residues and bind to proline-rich motifs devoid of aromatic residues. Group III WW domains encompass an additional residue in the first loop that is crucial for the recognition of phosphoSer/Thr-Pro motifs. Group IV WW domains were found to have the most divergent protein sequences of all WW domain groups; they lack the C-terminal Trp and their targets are so far unknown. In previous studies, group II WW domains have been shown to recognise SH3 domain ligands. To understand this phenomenon on a structural basis, a homology model of the Npw38 WW domain has been generated, which shows that not only the residues in the binding sites of the Sem5 SH3 and the Npw38 WW domain are very similar, but more importantly that the key residues in the binding sites can be perfectly superimposed. Despite the smaller size of the WW domain its binding pocket has in fact the same size as that of an SH3 domain.

Moreover, the solution structure of the tandem WW domains and the N-terminal FF domain of Prp40 have been determined by NMR (nuclear magnetic resonance) spectroscopy. To study the interaction of these Prp40 domains with the above-mentioned spliceosomal binding partners, ^1H - ^{15}N correlation spectra were recorded to map changes in the resonance frequencies of the domains upon ligand binding. It was found that the WW domains fold as triple-stranded anti-parallel β -sheets, which are not in direct contact, but separated by a

stable α -helical linker. Using ^1H - ^{15}N residual dipolar couplings (RDCs) the relative orientation of the WW domains has been precisely defined despite the lack of inter-domain distance restraints, revealing that the binding pockets face opposite sides. This supports the idea that the two Prp40 WW domains function as central adaptors in early spliceosomal complexes, bringing two different binding partners together. This suggestion has been confirmed by a detailed analysis of the proline-rich motifs recognised by the Prp40 WW domains. Both Prp40 WW domains are shown to interact with PPxY motifs present in the splicing factors BBP and Prp8, but also with related peptides devoid of aromatic residues. However, no interaction was observed between the Prp40 WW domains and the CTD repeats used questioning the direct function suggested for the Prp40 WW domains in coordinating splicing and transcription.

The N-terminal FF domain of Prp40 exhibits a novel α -helical fold. The three-dimensional structure of the Prp40 FF1 domain consists of four tightly packed α -helices. The second α -helix and its surrounding residues are shown to constitute the binding surface of this FF domain, which can accommodate the N-terminal TPR motif of the spliceosomal protein Clf1. These findings further cement the role of Prp40 as a central adaptor unit in the early spliceosome. Given that lethal U1 snRNA mutations can be suppressed by a point mutation in the second Prp40 FF domain (FF2), a homology model of the FF2 domain has been generated based on the determined FF1 domain structure. This model reveals significant differences in the electrostatic surface potential of the FF domain binding sites. In contrast to the FF1 domain, the binding surface of the FF2 domain is highly positively charged suggesting that the Prp40 FF2 domain can directly bind the U1 snRNA, although its sequence does not display any known RNA recognition motif.

In summary, the data presented in this Thesis shed new light onto the structural arrangement of splicing factors present in the yeast pre-spliceosome. The described results suggest the formation of a tight complex between the U1 snRNP, Prp40, BBP, Prp8 and Clf1 in the early spliceosome, which brings the 5' and the 3' splice-site into spatial proximity.

Zusammenfassung

Viele eukaryotische Proteine bestehen aus Domänen, deren Zusammenspiel die Funktion sogenannter Mosaikproteine oft entscheidend beeinflusst. Es ist daher von besonderem Interesse, nicht nur die Strukturen einzelner Proteindomänen zu kennen, sondern auch deren relative Anordnung. Der Spleißfaktor Prp40 (pre-mRNA processing protein), der in dieser Arbeit untersucht wurde, ist ein Mosaikprotein, das aus einem N-terminalen WW-Domänenpaar und sechs darauffolgenden FF-Domänen besteht. Im Hefe-Spleißosom bringt Prp40 die 3'- und 5'-Spleißstelle in räumliche Nähe, indem die WW-Domänen an Prolin-reiche Regionen der Spleißfaktoren BBP (branch-point binding protein) und Prp8 binden, während die FF-Domänen mit TPR-Motiven (tetratricopeptide repeats) des Spleißfaktors Clf1 (crooked neck-like factor 1) interagieren. Darüber hinaus vermutet man, daß die Wechselwirkung zwischen Prp40 und der phosphorylierten C-terminalen Domäne (CTD) der größten RNA Polymerase II-Untereinheit Transkription und mRNA-Spleißen zeitlich koppelt.

In der vorliegenden Arbeit wurde zunächst ein Versuch unternommen, die Mitglieder der WW-Domänenfamilie anhand ihrer Aminosäuresequenzen, dreidimensionalen Strukturen und Prolin-reichen Bindungsmotive zu klassifizieren. Dies führte zu einer Unterteilung von WW-Domänen in vier Gruppen. WW-Domänen der Gruppe I binden PPxY-Motive (wobei x ein beliebiger Aminosäurerest ist), während WW-Domänen der Gruppe II reich an aromatischen Resten sind und Prolin-reiche, aliphatische Motive erkennen. WW-Domänen der Gruppe III enthalten einen zusätzlichen Aminosäurerest im ersten Loop, der für die Erkennung von phosphoSer/Thr-Pro-Motiven verantwortlich ist. Gruppe IV WW-Domänen weisen die divergentesten Proteinsequenzen unter allen WW-Domänen auf; ihnen fehlt das C-terminale Trp, und bis heute sind keine Bindungspartner dieser Gruppe bekannt. Interessanterweise erkennen viele WW-Domänen der Gruppe II Liganden, die ursprünglich als Bindungspartner von SH3-Domänen bekannt waren. Dieses Phänomen konnte mittels eines Strukturmodells der Npw38 WW-Domäne erklärt werden. In diesem Modell passen die für die Ligandenerkennung verantwortlichen Aminosäurereste der Npw38 WW-Domäne strukturell perfekt mit denen der Sem5 SH3-Domäne übereinander. Obwohl WW-Domänen mit nur 35 Aminosäuren sehr viel kleiner sind als SH3-Domänen, konnte damit gezeigt werden, daß die Bindungstaschen dieser beiden Proteininteraktionsdomänen dennoch gleich groß sind.

Außerdem wurden die Strukturen des WW-Domänenpaares und der N-terminalen FF-Domäne des Spleißfaktors Prp40 mittels NMR-Spektroskopie (Kernspinresonanz oder nuclear magnetic resonance) gelöst. Mit Hilfe von ^1H - ^{15}N Korrelationsspektren wurde außerdem die Wechselwirkung dieser Prp40-Domänen mit den obengenannten Spleißfaktoren anhand der Änderung der Resonanzfrequenzen der Domänen bei Zugabe von Liganden studiert. Es konnte gezeigt werden, daß die WW-Domänen sich zu drei-strängigen anti-parallelen β -Faltblättern

falten, die durch einen stabilen α -helikalen Linker räumlich voneinander getrennt sind. Obwohl keine experimentellen Proton-Proton-Abstände zwischen den Domänen beobachtet wurden, konnte die relative Orientierung der WW-Domänen mit Hilfe residueller dipolarer ^1H - ^{15}N Kopplungen (RDCs) präzise bestimmt werden. Die relative Domänenanordnung zeigt, daß die Bindungstaschen der WW-Domänen in entgegengesetzte Richtungen zeigen. Dieses Ergebnis bestätigt die Vermutung, daß die Prp40 WW-Domänen als zentrale Adapter im frühen Hefe-Spleißosom fungieren und zwei verschiedene Bindungspartner in räumliche Nähe bringen können. Weitere Hinweise, die diese Vermutung stützen, erhielt man aus einer detaillierten Analyse Prolin-reicher Motive, die an die WW-Domänen des Prp40-Proteins binden. Diese Analyse zeigte, daß beide WW-Domänen die PPxY-Motive der Spleißfaktoren BBP und Prp8 binden. Außerdem wurden Wechselwirkungen zwischen den Prp40 WW-Domänen und verwandten Prolin-reichen Peptiden beobachtet, denen ein aromatischer Rest fehlte. Im Gegensatz dazu fand keine Wechselwirkung zwischen den Prp40 WW-Domänen und den verwendeten CTD-Sequenzen statt, was an einer direkten Funktion der Prp40 WW-Domänen an der Koordination von Transkription und Spleißen zweifeln lässt.

Die N-terminale FF(FF1)-Domäne des Prp40-Proteins weist eine neuartige α -helikale Faltung auf. Die dreidimensionale Struktur dieser FF-Domäne besteht aus vier dicht gepackten α -Helices. Die zweite α -Helix bildet zusammen mit den umgebenden Seitenketten die Bindungsfläche der FF-Domäne für das N-terminale TPR-Motiv des Spleißfaktors Clf1. Diese Ergebnisse untermauern die Rolle von Prp40 als zentrale Bindungsplattform im frühen Spleißosom. Da bekannt ist, daß eine Punktmutation in der zweiten Prp40 FF-Domäne (FF2) letale U1 snRNA-Mutationen unterdrückt, wurde ausgehend von der bereits bestimmten Struktur der FF1-Domäne ein Homologiemodell der FF-Domäne erstellt. Dieses Modell weist signifikante Unterschiede im elektrostatischen Oberflächenpotential der Bindungsstellen dieser beiden FF-Domänen auf. Im Gegensatz zur FF1-Domäne ist die Bindungsstelle der FF2-Domäne stark positiv geladen. Dies läßt vermuten, daß die FF2-Domäne direkt an U1 snRNA binden kann, obwohl ihre Aminosäuresequenz kein bisher bekanntes RNA-Erkennungsmotiv aufweist.

Abschließend kann man sagen, daß diese Arbeit neue Einblicke in die strukturelle Anordnung von Spleißfaktoren im frühen Hefe-Spleißosom liefert. Anhand der hier präsentierten Ergebnisse läßt sich ein Modell vorschlagen, in dem das U1 snRNP mit den Spleißfaktoren Prp40, BBP, Prp8 und Clf1 ein dichtes Protein-RNA-Netzwerk bildet, das dazu dient, die 3'- und 5'-Spleißstelle in räumliche Nähe zu bringen.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Pre-mRNA splicing | 3 |
| 1.1.1 | The splicing factor Prp40 | 3 |
| 1.2 | Nature's LEGO bricks: Protein domains | 6 |
| 1.2.1 | WW and FF domains | 6 |
| 1.3 | Aim of this Thesis | 8 |
| 1.4 | References | 9 |
| 2 | NMR spectroscopy | 13 |
| 2.1 | Principles of NMR spectroscopy | 13 |
| 2.1.1 | Effect of a static magnetic field | 14 |
| 2.1.2 | Radio frequency pulses | 15 |
| 2.1.3 | Chemical shift | 16 |
| 2.1.4 | Scalar and dipolar coupling | 16 |
| 2.2 | Quantum description of NMR | 17 |
| 2.2.1 | Product operators | 19 |
| 2.2.2 | Spin interaction Hamiltonians | 19 |
| 2.2.3 | External spin interactions | 20 |
| 2.2.4 | Internal spin interactions | 21 |
| 2.2.5 | Multi-dimensional NMR | 23 |
| 2.3 | NMR structure determination | 26 |
| 2.4 | References | 27 |
| 3 | Material and Methods | 29 |
| 3.1 | Material | 29 |
| 3.1.1 | Bacterial strains | 29 |
| 3.1.2 | Expression vectors | 29 |
| 3.1.3 | Buffers and solutions | 29 |
| 3.1.4 | Cell culture media | 30 |
| 3.1.5 | Recombinant proteins | 30 |
| 3.1.6 | Synthetic peptides | 32 |

| | | |
|----------|---|-----------|
| 3.2 | Biochemical methods | 32 |
| 3.2.1 | Transformation | 34 |
| 3.2.2 | Protein expression and purification | 34 |
| 3.2.3 | Determination of protein concentration | 35 |
| 3.3 | NMR spectroscopy | 35 |
| 3.3.1 | Resonance assignment | 36 |
| 3.3.2 | Derivation of structural restraints | 37 |
| 3.3.3 | Structure refinement and validation | 43 |
| 3.3.4 | Chemical shift mapping | 45 |
| 3.3.5 | Nuclear spin relaxation experiments | 45 |
| 3.4 | Other methods | 47 |
| 3.5 | References | 48 |
| 4 | Binding of Pro-rich ligands to WW and SH3 domains | 52 |
| 4.1 | WW domain as a phosphate-dependent SH3 domain? | 53 |
| 4.2 | What do we learn from the structures? | 54 |
| 4.2.1 | WW domains binding PPxY cores | 54 |
| 4.2.2 | The Pin1 and C-terminal repeat of the RNA polymerase II complex | 56 |
| 4.2.3 | Modeling of aromatic-rich WW sequences | 57 |
| 4.3 | Comparison between SH3 and WW binding sites: two different scaffolds for binding left-handed poly-proline type II helix ligands | 59 |
| 4.4 | Do we need extra interactions from neighboring domains to acquire specificity, to consolidate the WW domain structure or both? | 61 |
| 4.5 | References | 62 |
| 5 | Solution structure of the Prp40 WW domain pair | 66 |
| 5.1 | Resonance assignment and secondary structure | 67 |
| 5.2 | Structure determination and assessment | 70 |
| 5.3 | Description of the structure | 73 |
| 5.4 | Discussion and conclusions | 76 |
| 5.5 | References | 77 |
| 6 | Ligand recognition of the Prp40 WW domain pair | 81 |
| 6.1 | Interactions with proline-rich peptides | 82 |
| 6.2 | Interaction with PPxY motifs | 85 |
| 6.3 | Interaction with PPΨΨP motifs | 88 |
| 6.4 | Interaction with CTD RNA polymerase II peptides | 89 |
| 6.5 | Concluding remarks | 92 |
| 6.6 | References | 93 |

| | | |
|----------|---|------------|
| 7 | Structure and interactions of the Prp40 FF1 domain | 96 |
| 7.1 | Determination of the solution structure of the Prp40 FF1 domain | 98 |
| 7.1.1 | The Prp40 FF domain boundaries | 98 |
| 7.1.2 | Resonance assignment and description of the structure | 99 |
| 7.1.3 | Structure based sequence analysis of FF domains | 103 |
| 7.2 | NMR studies of the TPR1 repeat of the splicing factor Clf1 | 103 |
| 7.2.1 | Selection of the TPR repeats for interaction studies | 103 |
| 7.2.2 | Resonance assignment and secondary structure of the Clf1 TPR1 . . . | 104 |
| 7.3 | Interaction between the Prp40 FF1 domain and the TPR1 motif of Clf1 . . . | 105 |
| 7.3.1 | Identification of the Prp40 FF1 binding surface | 105 |
| 7.3.2 | Mapping of the Clf1 TPR1 binding site | 107 |
| 7.4 | Concluding remarks and discussion | 108 |
| 7.5 | References | 110 |
| | Abbreviations, symbols and units | 113 |
| | Acknowledgments | 117 |
| A | Experimental data | 119 |
| A.1 | Resonance assignments | 119 |
| A.2 | Coupling constants and dihedral angle restraints | 131 |
| A.3 | Residual dipolar couplings | 135 |
| A.4 | ¹⁵ N Relaxation data | 137 |