

# Crystal Structure of a Complex Between Human Spliceosomal Cyclophilin H and a U4/U6 snRNP-60K Peptide

Ulrich Reidt<sup>1</sup>, Markus C. Wahl<sup>1</sup>, Dirk Fasshauer<sup>2</sup>, David S. Horowitz<sup>3</sup>  
Reinhard Lührmann<sup>1</sup> and Ralf Ficner<sup>1,4\*</sup>

<sup>1</sup>Abteilung für Zelluläre Biochemie  
Max-Planck-Institut für Biophysikalische Chemie, 37077 Göttingen, Germany

<sup>2</sup>Abteilung für Neurobiologie  
Max-Planck-Institut für Biophysikalische Chemie, 37077 Göttingen, Germany

<sup>3</sup>Department of Biochemistry and Molecular Biology  
Uniformed Services University of the Health Sciences, Bethesda MD 20814, USA

<sup>4</sup>Abteilung für Molekulare Strukturbiologie, Institut für Mikrobiologie und Genetik  
Georg-August-Universität 37077 Göttingen, Germany

The spliceosomal cyclophilin H is a specific component of the human U4/U6 small nuclear ribonucleoprotein particle, interacting with homologous sequences in the proteins U4/U6-60K and hPrp18 during pre-mRNA splicing. We determined the crystal structure of the complex comprising cyclophilin H and the cognate domain of U4/U6-60K. The 31 amino acid fragment of U4/U6-60K is bound to a region remote from the cyclophilin active site. Residues Ile118–Phe121 of U4/U6-60K expand the central  $\beta$ -sheet of cyclophilin H and the side-chain of Phe121 inserts into a hydrophobic cavity. Concomitantly, in the crystal the cyclophilin H active site is occupied by the N terminus of a neighboring cyclophilin H molecule in a substrate-like manner, indicating the capacity of joint binding to a substrate and to U4/U6-60K. Free and complexed cyclophilin H have virtually identical conformations suggesting that the U4/U6-60K binding site is pre-shaped and the peptidyl-prolyl-*cis/trans* isomerase activity is unaffected by complex formation. The complex defines a novel protein–protein interaction mode for a cyclophilin, allowing cyclophilin H to mediate interactions between different proteins inside the spliceosome or to initiate from its binding platforms isomerization or chaperoning activities.

© 2003 Elsevier Ltd. All rights reserved

\*Corresponding author

**Keywords:** cyclophilin; peptidyl-prolyl *cis/trans* isomerase; spliceosome; snRNP; U4/U6-60K

## Introduction

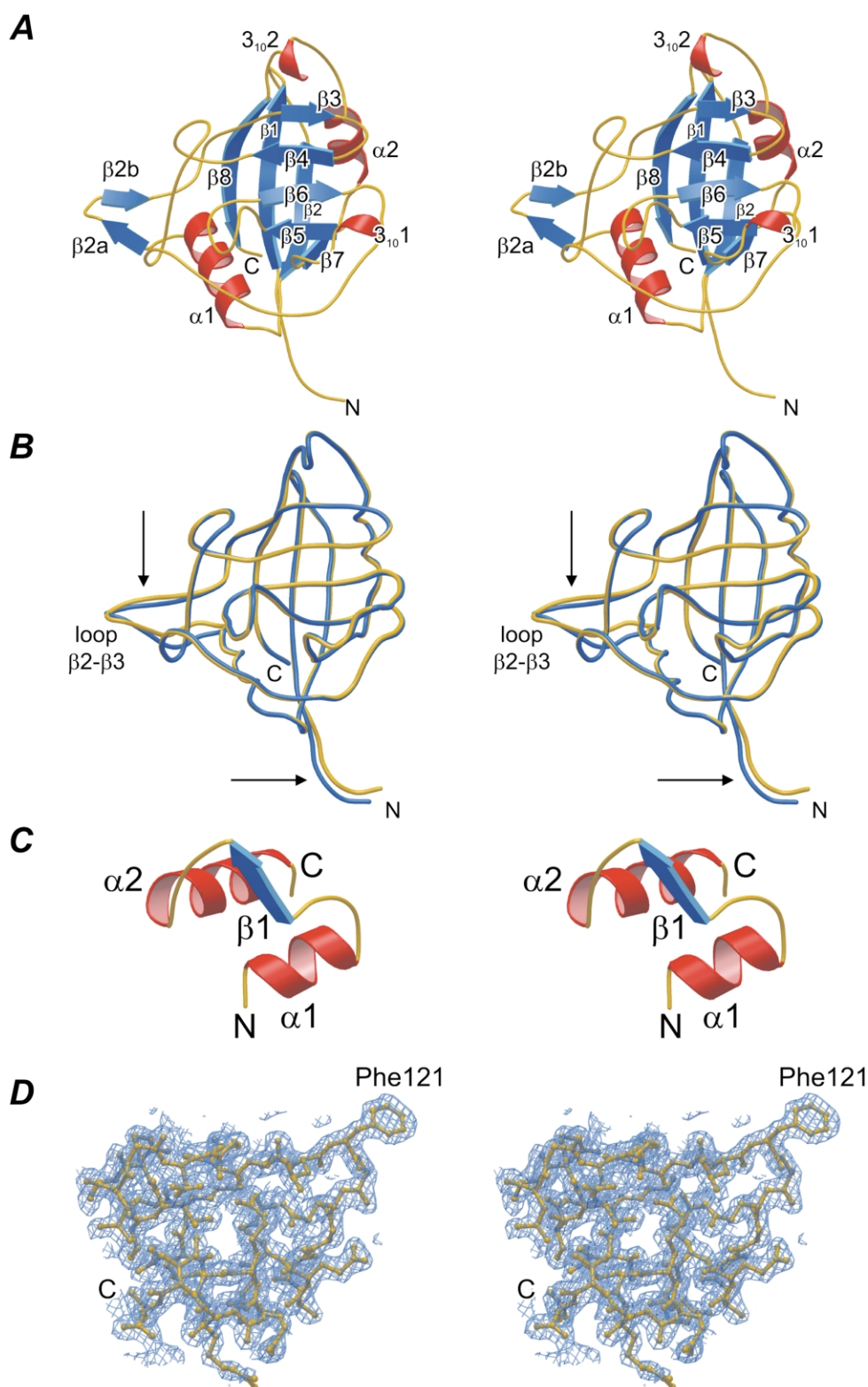
The expression of eukaryotic genes often requires the removal of non-coding, intervening sequences (introns) from precursor mRNA, a process catalyzed *via* two transesterification reactions by the spliceosome. The assembly of the spliceosome is a well-ordered process and involves snRNPs (small nuclear ribonucleoprotein particles) and non-snRNP protein factors.<sup>1</sup> In the early phase of spliceosome formation, U1 snRNP associates with the 5' splice site of the pre-mRNA and U2 snRNP interacts with the branch site, leading to the formation of the spliceosomal E and A com-

plexes, respectively.<sup>2,3</sup> The U4/U6 and U5 snRNPs assemble into a [U4/U6.U5] tri-snRNP before formation of the pre-spliceosome with the U1 and U2 particles.<sup>4,5</sup> Active spliceosomes are then formed through extensive RNA and protein conformational rearrangements, during which the U4/U6 duplex is disrupted and the U6 snRNA base-pairs with the U2 snRNA as well as the 5' splice site on the pre-mRNA. The new U2/U6 snRNA network is involved in the catalysis of the transesterification reactions.<sup>6</sup> The mechanisms underlying these structural rearrangements remain a mystery.

Each UsnRNP contains one (U1, U2, U5) or two (U4/U6) snRNAs and several proteins.<sup>7</sup> Common Sm proteins (B/B', D1, D2, D3, E, F, and G) are arranged around a highly conserved Sm site on the U1, U2, U4, and U5 snRNAs.<sup>8</sup> The remainder of the snRNP proteins is particle-specific.<sup>9</sup> Among the proteins unique to the U4/U6 snRNP is the 20 kDa cyclophilin H (CypH, SnuCyp-20, USA-Cyp) comprising 177 residues. CypH has so far

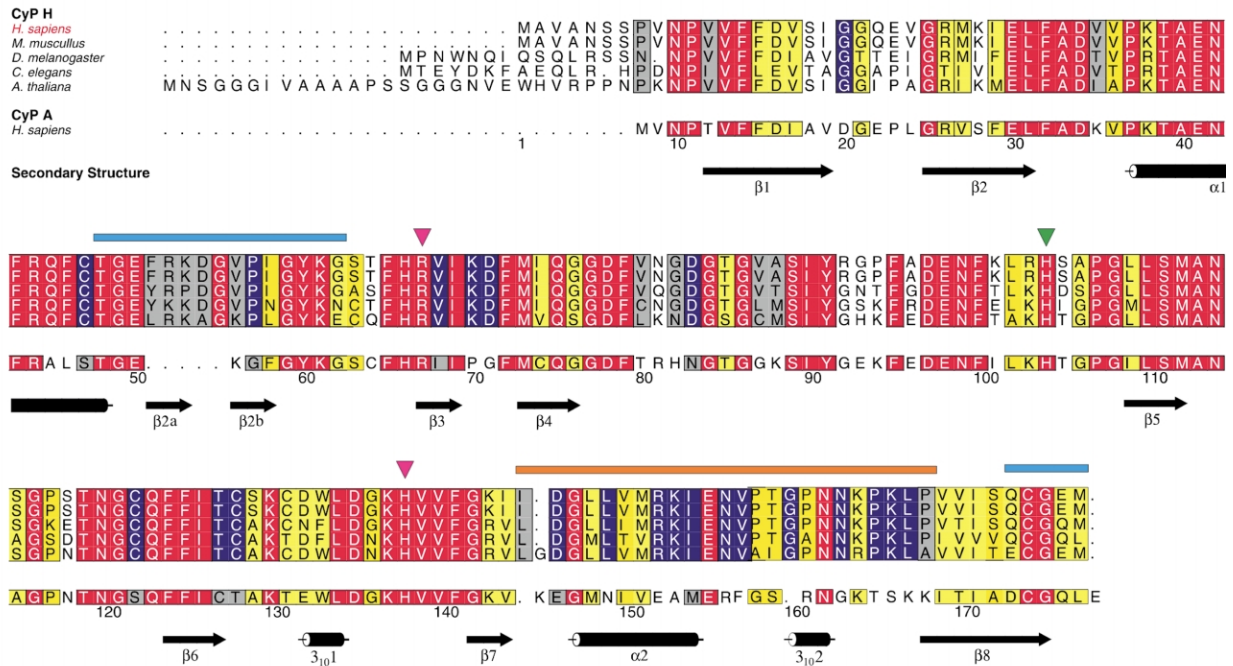
Abbreviations used: snRNPs, small nuclear ribonucleoprotein particles; GST, glutathione-S-transferase; hCypA, human cyclophilin A; PPIase, peptidyl-prolyl-*cis/trans* isomerase.

E-mail address of the corresponding author: rficner@gwdg.de

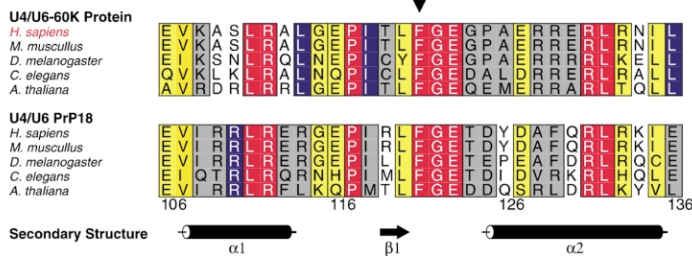


**Figure 1.** Overall structure. (A) Stereo ribbon plot portraying the overall fold of CypH in the present complex. Helices, red; strands, blue; coil, yellow. Secondary structure elements are labeled. If not stated otherwise, the following images are displayed from the same point of view. (B) Stereo plot of the superimposition of free CypH (yellow) and CypH as seen in the complex with the U4/U6-60K peptide (blue). Arrows mark regions with small local differences in the conformations. (C) Stereo ribbon plot of the structure of the 31-residue U4/U6-60K peptide as seen in the complex with CypH. (D) Final  $2F_o - F_c$  electron density at  $1\sigma$  encompassing the region of the U4/U6-60K peptide. If not otherwise mentioned, Figures were prepared with Molscript<sup>37</sup> and Bobscript (<http://www.strubi.ox.ac.uk/bobscript/>).

## A



## B



**Figure 2.** Sequence alignments. (A) Alignment of human and other putative CypH sequences and human CypA. Residues identical across the groups are on a red background, conserved residues are on a yellow background. Residues, which are only identical within the group of CypH-like cyclophilins are on a blue background, those conserved only within a subgroup are on gray. Secondary structure elements as seen in the present structure are indicated below the alignment. Regions binding the U4/U6-60K peptide are marked with a cyan bar above the alignment. Magenta arrows above the alignment point out the active site residues Arg67 and His138, contacting the substrate-like peptide through water-mediated hydrogen bonds. His104, which upon mutation yields an inactive enzyme but does not seem to contact the substrate directly, is indicated with a green arrow. A region in which the CypH-class and human CypA show significant sequence diversity, but which is remote from either of the identified functional sites on CypH in three dimensions, is indicated with an orange bar above the sequences. The protein used in the present study (human CypH) is labeled in red. *H. sapiens*, *Homo sapiens*; *M. musculus*, *Mus musculus*; *D. melanogaster*, *Drosophila melanogaster*; *C. elegans*, *Caenorhabditis elegans*; *A. thaliana*, *Arabidopsis thaliana*. (B) Alignment of regions corresponding to the U4/U6-60K peptide from the present structure. The first group aligns the corresponding regions from U4/U6-60K proteins, the second group aligns the homologous regions from Prp18 proteins. Observed secondary structure elements are indicated below the alignment. Phe121, which is most critical for binding to CypH, is labeled with a black arrow above the alignment. Color-coding is as in (A). The Figure was prepared with Alscript.<sup>38</sup>

only been identified in humans<sup>10</sup> and, in particular, an ortholog is lacking from the *Saccharomyces cerevisiae* proteome. Therefore, CypH may be an evolutionarily late addition to the arsenal of splicing factors.

The previously determined crystal structure of

CypH<sup>11</sup> shows the typical cyclophilin fold: an eight-stranded antiparallel  $\beta$ -barrel capped on both ends by an  $\alpha$ -helix (Figure 1(A)). The CypH structure is very similar to that of human cyclophilin A (hCypA). Consistent with the finding of peptidyl-prolyl-*cis/trans* isomerase (PPIase)

activities in hCypA<sup>12</sup> and CypH,<sup>10</sup> which can be inhibited by the immunosuppressive drug cyclosporin A (CsA),<sup>10,13</sup> their active sites exhibit almost identical structures.<sup>11</sup> The presence of an enzyme affecting the conformation of other protein components in the spliceosome may be relevant in light of the above described structural rearrangements during spliceosome maturation and catalysis.

CypH was shown to form a stable, RNA-free, ternary complex with the U4/U6 snRNP-specific 60K (U4/U6-60K) and 90K (U4/U6-90K) proteins. The latter two proteins are also referred to as hPrp4 and hPrp3, since they are the orthologues of the *S. cerevisiae* Prp4 and Prp3 splicing factors, respectively.<sup>14</sup> In the ternary complex CypH interacts with the N-terminal domain of U4/U6-60K,<sup>15</sup> while no direct contact of CypH with U4/U6-90K could be detected. In addition, CypH can interact with another spliceosomal protein, hPrp18.<sup>15</sup> A combination of anti-U4/U6-60K and anti-hPrp18 antibodies could largely deplete cellular extracts of CypH.<sup>15</sup> The only similar element in U4/U6-60K and hPrp18 proteins is a highly homologous region of 31 amino acids (Figure 2) and the corresponding peptides were shown to bind to CypH.<sup>15</sup> Furthermore, *in vitro* and *in vivo* splicing activities were inhibited, albeit not abolished, by CsA, and mutational analyses clearly showed that CypH was the target for the drug.<sup>15</sup> The data are consistent with CypH playing a subtle but decisive role both in the maturation and catalysis of the human spliceosome.

As a step towards elucidating the structural organization of the U4/U6 snRNP and the functional role of CypH during splicing, we have solved the co-crystal structure of CypH and the cognate 31 amino acid fragment of the U4/U6-60K protein. Our results clearly show that CypH recognizes the U4/U6-60K peptide in a highly specific manner through shape complementarity and electrostatic fit. Interestingly, the U4/U6-60K peptide binds to a predominantly hydrophobic cleft clearly distinct from the active site of CypH. In addition, we observe concomitant binding of the N terminus of a symmetry-related cyclophilin to the active site of CypH in a substrate-like manner. Thus, the current complex structure portrays an unprecedented cyclophilin–protein interaction motif, which leaves the enzyme catalytically competent for action on other proteins.

## Results

### Quality of the structure

CypH was overexpressed as a glutathione-S-transferase (GST) fusion in *Escherichia coli* and purified as described.<sup>11</sup> The cognate 31 amino acid fragment, corresponding to residues Glu106–Leu136 of the human U4/U6-60K protein (Figure 2), was chemically synthesized and HPLC-

**Table 1.** Data collection and refinement statistics

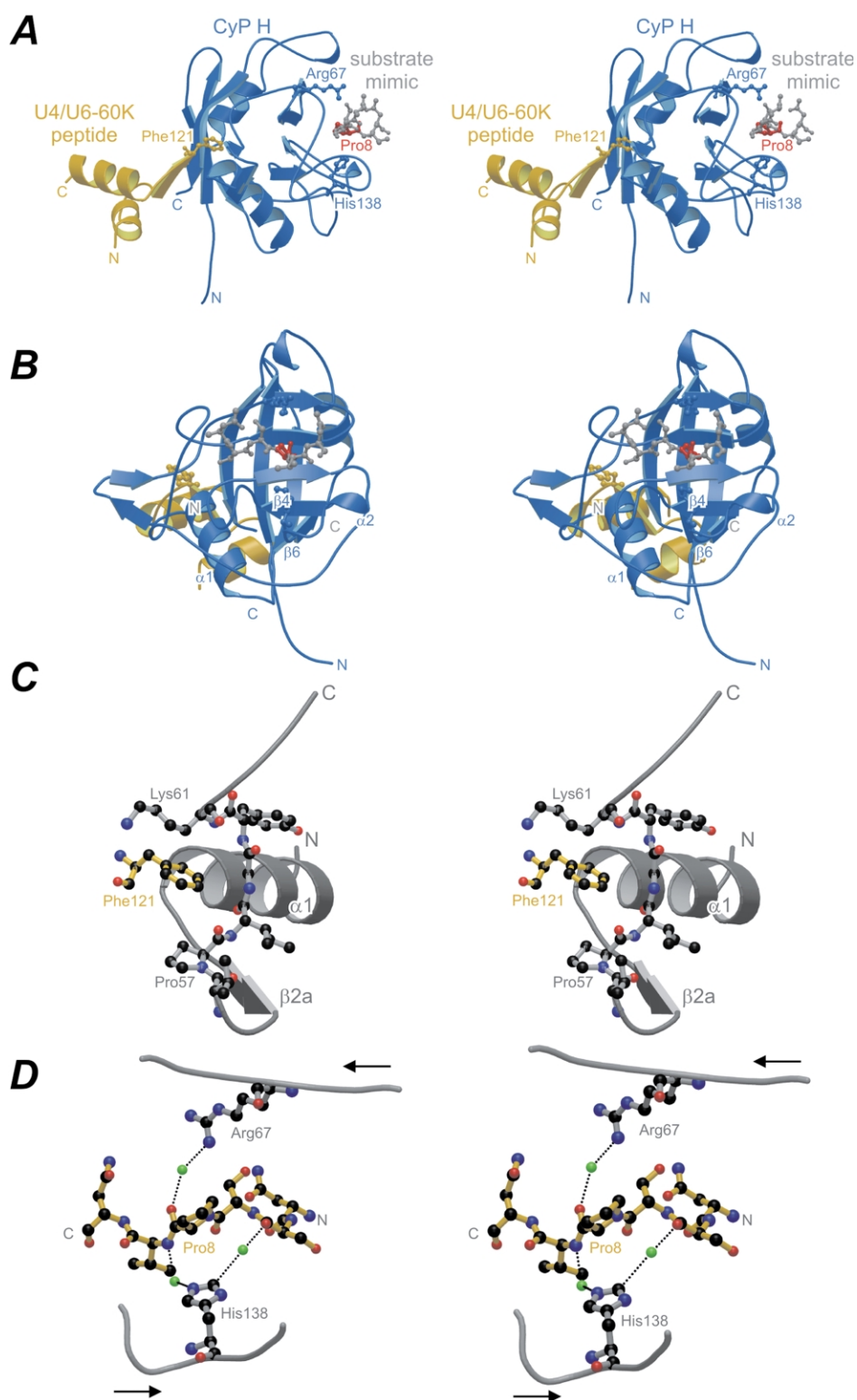
Data collection	
Space group	$P2_1$
Cell dimensions (Å)	$a = 44.4, b = 59.5, c = 47.1; \beta = 110.6^\circ$
Number of reflections	
Observed	55,492
Unique	13,762
Mosaicity (deg.)	1.5
Completeness (%)	
Overall (100.0–2.1 Å)	99.2
Last shell (2.1–2.1 Å)	99.0
$R_{\text{sym}}^a$ (%)	
Overall (100.0–2.1 Å)	6.3
Last shell (2.2–2.1 Å)	23.6
Refinement	
Resolution range (Å)	15.0–2.1
Number of reflections	13,458
R-factor <sup>b</sup> (%)	
Overall (15.0–2.1 Å)	19.5
Last shell (2.2–2.1 Å)	21.2
$R_{\text{free}}$ -factor <sup>b</sup> (%)	
Overall (15.0–2.1 Å)	25.7
Last shell (2.2–2.1 Å)	27.5
Number of residues in the final model	
CypH	173
U4/U6-60K peptide	31
Water oxygens	260
Average B-factors (Å <sup>2</sup> )	
CypH	12.4
U4/U6-60K peptide	15.8
Water oxygens	27.9
Rmsd from ideal geometry	
Bond lengths (Å)	0.005
Bond angles (deg.)	1.28
Ramachandran plot (%)	
Most preferred	82.0
Preferred	17.4
Additionally allowed	0.6
Disallowed	0

<sup>a</sup>  $R_{\text{sym}} = I - \langle I \rangle / \langle I \rangle$ , in which  $I$  is the observed intensity and  $\langle I \rangle$  is the average intensity for multiple measurements.

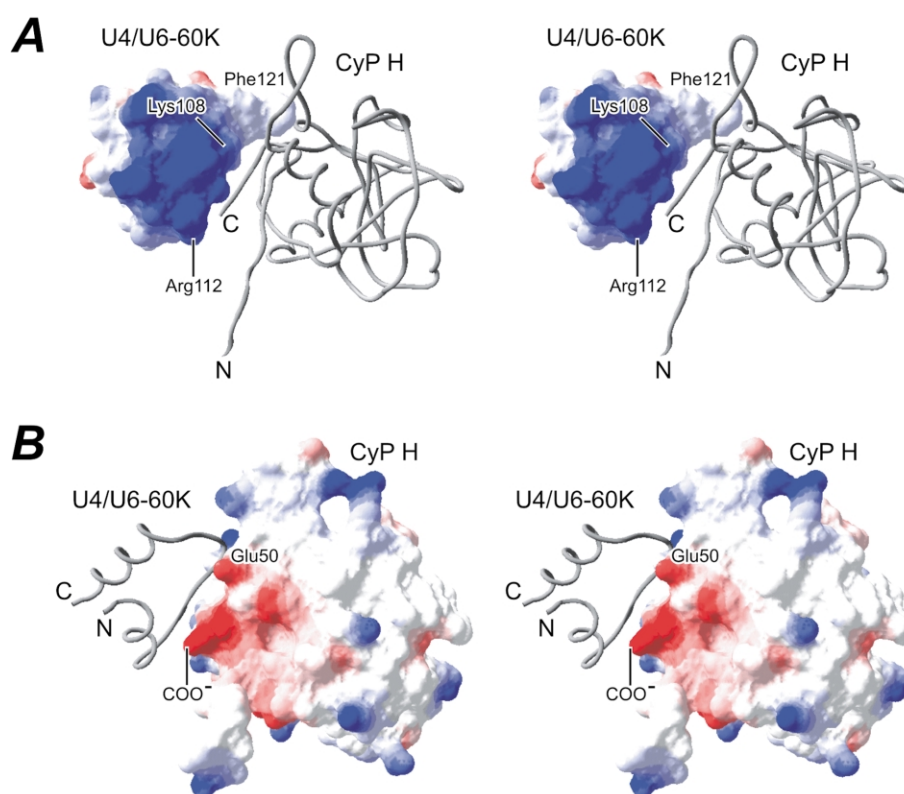
<sup>b</sup> R-factor =  $F_o - F_c / F_o$ , in which  $F_o$  and  $F_c$  are the observed and calculated structure factors, respectively. The  $R_{\text{free}}$ -factor is the cross-validation R-factor calculated for 10% of the reflections omitted in the refinement process.

purified. The crystal structure of the complex was solved by means of Patterson searches using the CypH structure<sup>11</sup> and refined at 2.1 Å resolution (Table 1). The final model comprises 208 amino acid residues and 425 water molecules. Its correctness is attested by a final overall R-factor and  $R_{\text{free}}$ -factor of 19.4% and 25.5%, respectively, for data between 15.0 Å and 2.1 Å resolution, combined with good stereochemistry. There were no outliers in a Ramachandran plot (Table 1)<sup>16</sup> and the final root-mean-square (rms) coordinate error estimated from a Luzatti analysis<sup>17</sup> was 0.23 Å. A  $2|F_o| - |F_c|$  map contoured at  $1\sigma$  showed no discontinuity in the electron density of the main chain and enveloped the vast majority of the side-chains.





**Figure 3.** Complex structure. (A), (B) Orthogonal stereo views of the CypH-U4/U6-60K peptide complex. CypH, blue; U4/U6-60K peptide, yellow. The N terminus of a neighboring cyclophilin (gray) is shown in ball-and-stick bound to the active site region of CypH. Pro8 of this peptide is shown in red. The orientation in (B) is the same as in Figure 1. (C) Close-up stereo plot of the binding pocket on CypH (gray) for Phe121 (yellow) of the U4/U6-60K peptide. Atoms are color-coded by atom type (C, black; N, blue; O, red). Residues and secondary structure elements cradling Phe121 of U4/U6-60K are labeled. (D) Close-up stereo plot of the active site region of one CypH molecule (gray) with active site residues Arg67 and His138 occupied by the N-terminal region of a neighboring molecule (yellow). The arrows indicate the directions of their pertaining  $\beta$ -strands ( $\beta 4$  and  $\beta 6$ , respectively). Pro8 of this substrate mimic is labeled. Water molecules bridging between Arg67/His138 and the active site ligand are shown in green with hydrogen bonds as broken lines.



**Figure 4.** Electrostatic surfaces. (A) and (B). Stereo views of the complex. The electrostatic potential is mapped on the surface of one component at a time with the other displayed in a ribbon representation. Positive potential, blue; negative potential, red. Residues involved in salt bridges are labeled at their positions on the surfaces. Relative to [Figure 1](#), the molecules were rotated 90° counterclockwise around the vertical axis and 45° into the plane of the paper around the horizontal axis. The Figure was produced with Swiss PDB viewer.<sup>39</sup>

### Binding of the U4/U6-60K peptide to a pre-shaped site on CypH

In the complex CypH exhibits the typical cyclophilin fold with eight antiparallel  $\beta$ -strands ( $\beta 1$ – $\beta 8$ ) curled into a central barrel and two peripheral  $\alpha$ -helices ([Figure 1\(A\)](#)). A surface loop between strands  $\beta 2$  and  $\beta 3$  ( $\beta 2$ – $\beta 3$  loop) exhibits an additional small  $\beta$ -hairpin (strands  $\beta 2a$  and  $\beta 2b$ ). The U4/U6-60K protein fragment folds into two  $\alpha$ -helices ( $\alpha 1$ : Val107–Ala113;  $\alpha 2$ : Pro125–Ile135; numbering is according to the full-length U4/U6-60K protein) connected by an extended loop (Leu114–Gly124; [Figure 1\(C\)](#)). Secondary structure elements for both components are identified in [Figure 1](#).

The total surface area buried upon contact formation is 878 Å<sup>2</sup> (both partners). U4/U6-60K residues Thr119 and Leu120 of the extended loop adopt a short  $\beta$ -strand conformation and, together with the neighboring residues, bind in antiparallel fashion to strand  $\beta 8$  of CypH, thus extending the cyclophilin  $\beta$ -sheet by one strand ([Figure 3\(A\)](#) and [\(B\)](#)). The arrangement leads to the formation of backbone-to-backbone hydrogen bonds between Cys174, Gly175, and Glu176 of CypH and Phe121, Thr119, and Ile118 of the peptide. In addition, there is a tight contact between loop  $\beta 1$ – $\alpha 2$  of the

peptide and loop  $\beta 2$ – $\beta 3$  of CypH, in which Lys108 (U4/U6-60K) and the backbone nitrogen of Gly122 (U4/U6-60K) form a salt bridge to Glu50 (CypH) and a hydrogen bond to the backbone carbonyl of Thr48 (CypH), respectively ([Figure 4](#)). An additional weaker salt bridge forms between Arg112 (U4/U6-60K) and the terminal carboxyl group of CypH ([Figure 4](#)). As a result of the above alignment, the side-chain of Phe121 from the U4/U6-60K peptide is nestled in a hydrophobic pocket of CypH ([Figure 4\(A\)](#)), mainly formed by CypH residues Pro57, Ile58, Gly59, Tyr60, and Lys61. The phenylalanine ring is sandwiched between Pro57 and the aliphatic portion of the Lys61 side-chain. The C-terminal turn of helix  $\alpha 1$  of CypH forms the floor of the pocket ([Figure 3\(C\)](#)). There are no bridging water molecules found in the interface. The two components are thus aligned for a snug fit, in which hydrophobic interactions contribute shape complementarity, salt bridges supply an electrostatic fit, and hydrogen bonds additionally fix the directionality of the interaction.

The calculated rms deviation (rmsd) for 171 common C $\alpha$  atoms of the free and complexed CypH is 0.56 Å and there are no large local differences. Upon overlaying the two molecules ([Figure 1\(B\)](#)), small deviations are seen in the N termini and in the  $\beta 2$ – $\beta 3$  loops, both of which correspond to presumably

flexible surface elements, which are stabilized by lattice interactions. Therefore, the structure of CypH is virtually identical in the two forms, indicating that the U4/U6-60K binding site is pre-shaped in the free enzyme.

We recorded circular dichroism (CD) spectra of the U4/U6-60K peptide ( $10^{-5}$  M) at 25 °C and 4 °C in physiological salt solution. Not surprisingly and in contrast to CypH, the peptide largely adopts a random coil conformation (data not shown). At 4 °C a small fraction of  $\alpha$ -helix can be detected (less than 10%). Therefore, the entropic cost of folding upon interaction with CypH is made up by the binding energy, suggesting that the complex observed herein is strong and specific. It remains an open question whether the corresponding region in the full-length U4/U6-60K protein is unfolded or maintains a fold similar to that seen in the complex with CypH.

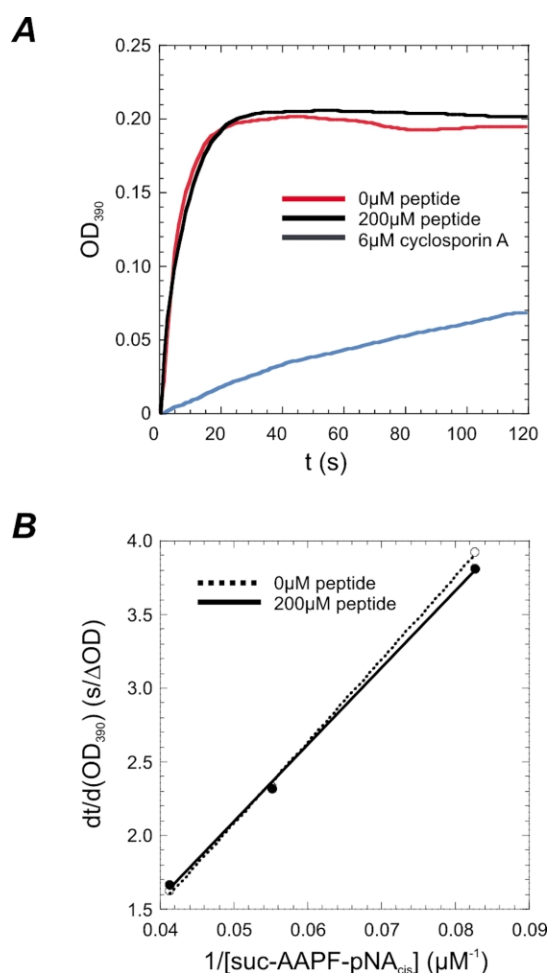
### Phylogenetic comparisons

Generally, the cross-species conservation of residues mediating critical interactions inside a protein or between interaction partners is an indicator for the specificity and ubiquity of the observed structure. CypH has so far only been positively identified in humans as a binding partner to U4/U6-60K and hPrp18. An open reading frame (98% identical on the protein level) can be discerned in the mouse genome, suggesting that at least mammals share CypH proteins. In addition, there are close sequence homologues in *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Arabidopsis thaliana* (around 80% amino acid identity) (Figure 2). In the above five cyclophilins, all residues contacting the U4/U6-60K Phe121 in the human system are absolutely conserved (Figure 2), which is not the case, e.g. with human CypA or CypB. Furthermore, the cyclophilins potentially belonging to the CypH class all show a conserved insertion of five residues between the C terminus of helix  $\alpha$ 1 and the start of the residues lining the hydrophobic pocket for U4/U6-60K Phe121. This insertion is known as the divergent loop<sup>18</sup> and represents one of two sequence additions, besides another eight residues extending the N terminus, emerging upon comparison to related cyclophilins (Figure 2). While ultimately it has to be shown directly that these CypH-like cyclophilins are associated with their respective spliceosomes, the above sequence comparisons strongly suggest such a function.

In addition, members of the U4/U6-60K and Prp18 protein families from the above organisms, which potentially encode CypH orthologues, also show significant sequence identities across the region that is involved in CypH binding in the human systems. Conspicuously, the U4/U6-60K and Prp18 orthologues from *S. cerevisiae*, which apparently lacks an orthologue of CypH, do not (Figure 2). Hence a comparison of the U4/U6-60K and Prp18 orthologues from the organisms with

putative CypH can identify residues that are critical for the interactions with the cyclophilin. In addition, the comparison between the CypH-binding sites in human U4/U6-60K and hPrp18 should also hint at critical residues (Figure 2).

Among the relevant U4/U6-60K proteins, the residues mediating the salt bridges to CypH are absolutely conserved (Arg112) or maintain the positive charge (Lys108). Furthermore, Phe121, the main determinant for the shape complementarity, is conserved throughout all the sequences, as is the following Gly122. The latter residue seems to be important for the  $\beta$ 1- $\alpha$ 2 loop conformation in U4/U6-60K, which allows its close approach to



**Figure 5.** CypH PPIase activity in presence of U4/U6-60K peptide. (A) Progress of the isomerization of the substrate peptide (12.1  $\mu$ M *cis* concentration), catalyzed by CypH in the absence (red) or presence (200  $\mu$ M; black) of the U4/U6-60K peptide. A trace of a reaction with 6  $\mu$ M cyclosporin A (CsA) instead of U4/U6-60K peptide under otherwise identical conditions is shown for comparison (blue). (B) Double-reciprocal plot of PPIase activity at various substrate concentrations (*cis* form) in the absence (broken line) or presence (200  $\mu$ M; continuous line) of the U4/U6-60K peptide. Data points represent averages of at least two independent measurements.

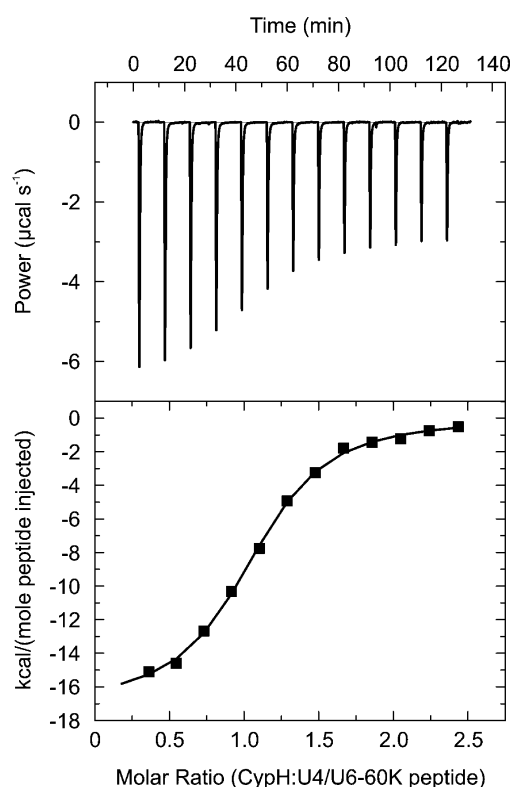
the backbone of the CypH  $\beta 2$ - $\beta 3$  loop. The remainder of the conserved residues in the U4/U6-60K peptides is important for the structural maintenance: Glu123 engages in a salt bridge with Arg131; Arg128 forms a hydrogen bond to the side-chain of Thr119; Val107, Leu111, and Leu132 of U4/U6-60K point to the interior of the helix-loop-helix motif and stabilize its present conformation through hydrophobic interactions.

Significantly, the side-chains of Val107, Leu111, Arg112, Phe121, Glu122, Arg131, and Leu132 in U4/U6-60K, which maintain important inter- or intramolecular contacts, are also conserved in the corresponding regions of Prp18 proteins (Figure 2). Furthermore, the structurally important Pro117 and Gly122 are absolutely conserved across the groups of U4/U6-60K and Prp18 proteins. Such

conservation suggests that the homologous peptide from Prp18 molecules should be able to adopt the same structure and interact in largely the same manner as described herein for the U4/U6-60K fragment.

The peptide used for co-crystallization was originally identified by sequence alignments.<sup>15</sup> It may represent the minimal binding sequence of U4/U6-60K and Prp18 proteins for CypH. Additional contacts between the interaction partners may exist when full-length proteins are employed.

The present model of binding to a ligand peptide is unprecedented in other known cyclophilin-protein interactions. In the case of hCypA, the enzyme was shown to bind *via* its active site to the area of Pro90 in an exposed loop of the HIV-1 capsid protein.<sup>19</sup>



**Figure 6.** Thermodynamics of the interaction of CypH and the U4/U6-60K peptide. Isothermal titration calorimetric analysis of the interaction of CypH (28.9  $\mu\text{M}$ ) with the U4/U6-60K peptide (900  $\mu\text{M}$ ) at 25  $^{\circ}\text{C}$ . Raw data were obtained over a series of injections of U4/U6-60K peptide and plotted as power ( $\mu\text{cal}/\text{seconds}$ ) versus time (minutes). Upper panel: The negative peaks (black line) indicate an exothermic reaction with the peak area proportional to the heat released at each injection (red line: base line). Lower panel: Binding isotherms created by plotting the areas under the peaks in the upper panel against the molar ratio of the U4/U6-60K peptide injected to CypH. The best-fit line through the values was obtained by least-squares regression using a one-site model and yields the enthalpy ( $\Delta H_a$ ), stoichiometry ( $n$ ), and the equilibrium dissociation constant ( $K_d$ ). For the interaction of CypH with the U4/U6-60K peptide  $n = 1.022$ ,  $\Delta H_a = -17.08 \text{ kcal mol}^{-1}$ , and  $K_d = 1.97 \mu\text{M}$ .

### Binding of a substrate-like peptide

CypH harbors a proline residue (Pro8) in its N-terminal region, which is extended in comparison to hCypA and other cyclophilins (Figure 2). Fortunately, in the crystalline lattice, the sequence encompassing Pro8 of one CypH molecule inserts into the active site along  $\beta$ -strands  $\beta 4$  and  $\beta 6$  of a neighboring cyclophilin (Figure 3(D)). 738  $\text{\AA}^2$  of surface area from both molecules are buried upon contact formation. The N-terminal sequence of CypH in the crystal thus seems to constitute a substrate mimic. The present packing is similar to the arrangement seen in the structure of free CypH (our unpublished observations).

The Ser7-Pro8 peptide bond, which could undergo isomerization, is presently seen in a *trans* configuration. The region around Pro8 in the active site ligand is contacted through water-mediated hydrogen bonds by the side chains of Arg67 and His138 (Figure 3(D)). In the present model, NH1 of Arg67 is hydrogen-bonded to water W21, which in turn is in contact with the carbonyl oxygen of Pro8. Two additional water molecules, W44 and W48, are in hydrogen bonding distances to CE1 and NE2 of His138, respectively. Due to this crystal packing interaction, the active center of CypH is not available for ligands.

### PPIase activity of CypH-U4/U6-60K peptide complex

Since the U4/U6-60K peptide contains two proline residues (Pro117, Pro125), the question arises whether the U4/U6-60K peptide could bind to the active site of CypH and therefore could act as a competitive inhibitor. PPIase activity tests were carried out in the presence and absence of U4/U6-60K peptide using CypH and varying concentrations of the substrate (suc-AAPF-pNA) (Figure 5). The inspection of the double reciprocal plots unambiguously shows that the U4/U6-60K peptide does not inhibit the CypH PPIase activity.



This confirms that the U4/U6-60K peptide binding site is different from the catalytic center.

### Binding constant of U4/U6-60K peptide and stoichiometry of the complex in solution

The enthalpy of binding of the U4/U6-60K peptide to CypH was measured by isothermal titration calorimetry revealing an exothermic reaction (Figure 6). The equilibrium dissociation constant ( $K_d$ ) was determined to 1.97  $\mu$ M. Fitting of the binding isotherm indicated a stoichiometry for the association ( $n$ ) of 1.022 (peptide: CypH) with no secondary binding. This number is consistent with a single peptide binding site on CypH. Since the PPIase inhibition studies (see above) excluded binding of the peptide to the active site of CypH, the experimentally determined binding stoichiometry confirms the novel binding site observed in the crystal structure.

## Discussion

### A novel cyclophilin–protein interaction site generates two business ends without crosstalk

We have determined the co-crystal structure of CypH in complex with a cognate peptide of the U4/U6-60K protein, exhibiting the minimal binding site on the latter protein. The structure defines at molecular detail the interaction between CypH and two other spliceosomal protein components, generating restraints for the structures of the CypH–U4/U6-60K and the CypH–hPrp18 complexes. The CypH active site and the interaction site for U4/U6-60K or Prp18 proteins are located on opposite sides of the enzyme. All other known cyclophilin–protein interactions, e.g. the human CypA–HIV-1 capsid complex,<sup>19</sup> occur exclusively through the cyclophilin active site. Hence this complex structure outlines unexpected protein–protein interaction capacities for the cyclophilin fold.

Significant affinity of the peptide to CypH in solution was demonstrated by ITC revealing a stoichiometric interaction. The PPIase activity tests of CypH in presence of the peptide demonstrated that CypH does not bind the peptide *via* its active site. Furthermore, mutations in the divergent loop region  $\alpha 1$ – $\beta 3$ , which is part of the peptide binding site, abolish binding of the U4/U6-60K protein (D. Ingelfinger *et al.*, unpublished results). These results confirm the physiological role of the binding site observed in the complex crystal structure.

The substrate-like interaction of neighboring CypH molecules in the crystal directly demonstrates the feasibility of concomitant binding of CypH to a substrate and to the U4/U6-60K or hPrp18 protein. In agreement with this notion it was observed that the PPIase activity of CypH is unaffected by incorporation into the [U4/U6.U5] tri-snRNP<sup>10</sup> and that mutations in the active site of CypH abolish its catalytic activity without affecting

U4/U6-60K binding (D. Ingelfinger *et al.*, unpublished results). Together, the data firmly establish the local and functional separation of the two interaction sites on CypH.

### Spliceosomal sub-assemblies centered on U4/U6-60K and Prp18

CypH is part of a stable hetero-trimeric substructure with the U4/U6-60K and U4/U6-90K proteins in the pre-spliceosome.<sup>10</sup> While the U4/U6-60K protein was identified as a direct binding partner of CypH,<sup>10</sup> no direct interaction between CypH and the U4/U6-90K has so far been observed. How are the proteins assembled into a larger network?

U4/U6-60K and its orthologue Prp4 contain a C-terminal WD40 domain with seven WD40 repeats, presumably folded into a seven-bladed  $\beta$ -propeller as seen in the  $\beta$ -subunits of G-proteins.<sup>20,21</sup> The alleged large flat surface of the propeller is a suitable interface for protein–protein interactions<sup>20</sup> in accord with mutational and two-hybrid studies for the Prp4–Prp3 complex.<sup>14</sup> In contrast, the U4/U6-60K–CypH interactions take place through the N-terminal part of U4/U6-60K, which does not belong to the  $\beta$ -propeller. Therefore, U4/U6-60K contains separate binding domains for U4/U6-90K and CypH and forms the central component of the tripartite complex, while U4/U6-90K and CypH do not interact directly or interact only transiently. U4/U6-60K leaves this complex together with the U4 snRNA upon conversion of the pre-spliceosome to the active spliceosome.<sup>22</sup>

CypH also interacts with Prp18, which is important for the second transesterification step in splicing.<sup>15</sup> Several functional domains have been defined in Prp18. While the N-terminal region of Prp18 contains the U4/U6-60K-like peptide for interaction with CypH,<sup>15</sup> the C-terminal domain, whose crystal structure has recently been determined, seems to harbor additional interaction motifs for the spliceosomal protein Slu7 and for the U5 snRNP, both of which seem to interact with 3' splice sites.<sup>23</sup> Therefore, also at later stages of the splicing reaction, CypH appears to be indirectly connected to other components of the spliceosome through Prp18.

Taken together, U4/U6-60K and Prp18 seem to constitute the platforms from which CypH initiates its activities, either the binding to other spliceosomal components, or chaperone and isomerization functions at several points during splicing. One or several of these CypH activities seem to be useful both at early stages of spliceosome assembly (U4/U6-60K interaction) and during later catalytic steps (Prp18 interaction). Alternatively, U4/U6-60K and Prp18 could use the cyclophilin as a bridge to yet other components of the particles. Interestingly, another region between Ile145 and Pro168 of the CypH-like cyclophilins can be discerned, which differs significantly in sequence

from CypA (Figure 2, orange bar). This region is remote from either the active site or the U4/U6-60K peptide binding site and could mark another contact region to yet another spliceosomal component.

### Implications for the cellular deployment of the cyclophilin fold

Since cyclophilins are ubiquitously expressed and have been conserved from bacteria to mammals, they must be important components of the cellular inventory. However, it has been difficult to assign precise functional roles to specific members of this uniformly folded but seemingly diversely employed family of enzymes. As seen with Nina A from *Drosophila*, which binds to and thus prevents misfolding of rhodopsin,<sup>24</sup> cyclophilins may act as chaperones. Other studies have shown that cyclophilins, including CypH, can catalyze the isomerization of prolyl-peptide bonds, a process which can be rate limiting in protein folding.<sup>25,26</sup> Together, the observations of specific interactions of CypH with known components of the spliceosome, the central importance of structural dynamics in the pre-spliceosome, and the PPIase activity of CypH suggest that it could mediate specific folding processes within the spliceosomal particles. Interestingly, further cyclophilins have recently been identified as novel components of the spliceosome.<sup>27</sup>

Given the similar overall fold of cyclophilins, their involvement in a particular cellular process requires the evolution of specific protein interaction sites. Significantly, the present structure defines a new way in which a cyclophilin can interact with another protein. In contrast to previously known contact modes, the present complex leaves the active site of the enzyme untouched so that it remains available for catalysis. Therefore, cyclophilins display a remarkable interaction versatility. This versatility presumably led the spliceosome to recruit a member of the cyclophilin family to its core, comparable to the employment of the thioredoxin fold in the spliceosomal 15K protein.<sup>28</sup>

## Materials and Methods

### Complex production, crystallization and data collection

CypH was overexpressed and purified as described.<sup>11</sup> The purified protein was stored in 120 mM NaCl, 2 mM DTT, 20 mM Tris-HCl, pH 7.5 (buffer A) and concentrated to 34 mg/ml by ultrafiltration. A 31 amino acid peptide corresponding to the suggested CypH binding site on human U4/U6-60K<sup>15</sup> was synthesized. The lyophilized peptide was dissolved to a concentration of 10 mg/ml in buffer A. For complex formation the two components were mixed in a 1:1 volumetric ratio. Co-crystals of the complex were grown at 21 °C in Linbro plates by the hanging drop vapor diffusion technique. Screening of sparse matrix buffer formulations yielded

preliminary crystallization conditions. Subsequent refinement optimized the well buffer to 2% (w/v) PEG8000, 200 mM magnesium acetate, 0.1 M Hepes, pH 6.5. Crystals formed within one week and could be flash-frozen in a liquid nitrogen stream after transfer into Paratone-N. X-ray data were collected on a Nonius FR591 rotating anode generator, operating at 50 kV and 100 mA, equipped with a MAR-345 image plate detector and Osmic focusing mirrors. 1° oscillation images were collected with a 900 seconds exposure time and a crystal-to-detector distance of 120 mm (Table 1). Diffraction data were processed using the programs DENZO and SCALEPACK.<sup>29</sup> Density considerations suggested one complex per crystallographic asymmetric unit.

### Structure solution and refinement

The structure of the complex was solved by molecular replacement as implemented in CNS,<sup>30</sup> using the crystal structure of CypH as a search model.<sup>11</sup> The cross-rotation function of CNS resulted in two peaks significantly above the background. The translation search on the second peak of the cross-rotation function yielded a solution with a correlation coefficient of 60.2% and a *R*-factor 38.4% for data between 15.0 Å and 2.1 Å. This solution was further refined in CNS. A round of rigid body refinement was followed by refinement of atomic positions and *B*-factors. The initial electron density map was significantly improved with the help of the program wARP.<sup>31</sup> This improved density clearly revealed a section, which was not accounted for by the CypH model and was readily interpretable in terms of the U4/U6-60K peptide used in the co-crystallization. With the exception of Glu106, which was built as alanine, a complete model for the peptide was built into the electron density map with the program O.<sup>32</sup> The structure of the CypH comprising residues 5–177 was manually checked, and the complex structure was further refined with CNS and wARP. In subsequent rounds of refinement, water molecules were built with CNS into vacant peaks of the  $|F_o| - |F_c|$  and  $2|F_o| - |F_c|$  electron density maps. The final model contains 425 water molecules and has good stereochemistry as evaluated with PROCHECK.<sup>33</sup>

### CD spectroscopy

CD spectra of the U4/U6-60K peptide were recorded with a J-715 spectropolarimeter (JASCO Corp., Tokyo, Japan). The analyte was dissolved at 10<sup>-5</sup> M in 150 mM NaCl, 20 mM Tris-HCl, pH 7.0 and the data acquired at 25 °C and 4 °C. Secondary structure predictions were done with the program CDNN.<sup>34</sup>

### Assay for PPIase activity

CypH PPIase activity was determined using the synthetic peptide substrate, *N*-succinyl-Ala-Ala-Phe-*p*-nitroanilide (suc-AAPF-*p*NA), in a chymotrypsin-coupled colorimetric assay.<sup>12</sup> At equilibrium, about 12.1% of the substrate is in the *cis* conformation.<sup>35</sup> Assays were performed using a mixture containing 35 mM Hepes, pH 7.9, 35 nM CypH and varying concentrations of suc-AAPF-*p*NA either in the absence or presence (200 μM) of the U4/U6-60K peptide. As a control, reactions were also monitored in the presence of 6 μM CsA. The mixtures were allowed to equilibrate at 8 °C for ten minutes before initiating the reactions with 5 mg

of  $\alpha$ -chymotrypsin. The reactions were followed through absorbance changes at 390 nm at two seconds intervals for three minutes using a Ultraspec3000pro spectrophotometer (Amersham Pharmacia).

### Isothermal titration calorimetry (ITC)

ITC measurements were carried out at 25 °C using a MCS-ITC Instrument (MicroCal, Northampton, MA) to obtain enthalpy and heat capacity changes.<sup>36</sup> Samples were dialyzed against degassed buffer A. The U4/U6-60K peptide at a concentration of 900  $\mu$ M was injected into the reaction cell (1.37 ml) containing CypH at a concentration of 29  $\mu$ M. A 100  $\mu$ l syringe was used to give a series of 8  $\mu$ l injections at 600 seconds intervals. The data were analyzed with the software Microcal Origin 5.0, yielding the enthalpy of binding ( $\Delta H_a$ ), the equilibrium dissociation constant ( $K_d$ ), and the stoichiometry ( $n$ ).

### Protein Data Bank accession numbers

The coordinates of the co-crystal structure have been deposited with the Protein Data Bank under accession number 1MZW and will be released upon publication.

### Acknowledgements

We are grateful to M. Krause (Institut für Molekularbiologie und Tumorforschung, Philipps Universität Marburg), who synthesized and purified the U4/U6-60K peptide, and to our colleagues D. Ingelfinger and T. Achsel for providing results prior to publication. R.F. was supported by the Deutsche Forschungsgemeinschaft (SFB286, TPA11).

### References

- Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* **65**, 367–409.
- Michaud, S. & Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.* **5**, 2534–2546.
- Michaud, S. & Reed, R. (1993). A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev.* **7**, 1008–1020.
- Will, C. L. & Lührmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290–301.
- Will, C. L. & Lührmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **9**, 320–328.
- Lamond, A. I., Sproat, B., Ryder, U. & Hamm, J. (1989). Probing the structure and function of U2 snRNP with antisense oligonucleotides made of 2'-OMe RNA. *Cell*, **58**, 383–390.
- Reddy, R. (1986). Compilation of small RNA sequences. *Nucl. Acids Res.* **14**, 61–72.
- Raker, V. A., Plessel, G. & Lührmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle *in vitro*. *EMBO J.* **15**, 2256–2269.
- Makarova, O. V., Makarov, E. M. & Lührmann, R. (2001). The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *EMBO J.* **20**, 2553–2563.
- Teigelkamp, S., Achsel, T., Mundt, C., Gothel, S. F., Cronshagen, U., Lane, W. S., *et al.* (1998). The 20 kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60 and 90 kD proteins. *RNA*, **4**, 127–141.
- Reidt, U., Reuter, K., Achsel, T., Ingelfinger, D., Lührmann, R. & Ficner, R. (2000). Crystal structure of the human U4/U6 small nuclear ribonucleoprotein particle-specific SnuCyp-20, a nuclear cyclophilin. *J. Biol. Chem.* **275**, 7439–7442.
- Fischer, G., Bang, H. & Mech, C. (1984). Determination of enzymatic catalysis for the *cis-trans* isomerization of peptide binding in proline-containing peptides. *Biomed. Biochim. Acta*, **43**, 1101–1111.
- Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T. & Schmid, F. X. (1989). Cyclophilin and peptidyl-prolyl *cis-trans* isomerase are probably identical proteins. *Nature*, **337**, 476–478.
- Ayadi, L., Callebaut, I., Saguez, C., Villa, T., Mornon, J. P. & Banroques, J. (1998). Functional and structural characterization of the prp3 binding domain of the yeast prp4 splicing factor. *J. Mol. Biol.* **284**, 673–687.
- Horowitz, D. S., Lee, E. J., Mabon, S. A. & Misteli, T. (2002). A cyclophilin functions in pre-mRNA splicing. *EMBO J.* **21**, 470–480.
- Ramachandran, G. N. S. V. (1968). Conformation of polypeptides and proteins. *Advan. Protein Chem.* **23**, 283–437.
- Luzzati, V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallog. sect. A*, **5**, 802–810.
- Dornan, J., Page, A. P., Taylor, P., Wu, S., Winter, A. D., Husi, H. & Walkinshaw, M. D. (1999). Biochemical and structural characterization of a divergent loop cyclophilin from *Caenorhabditis elegans*. *J. Biol. Chem.* **274**, 34877–34883.
- Gamble, T. R., Vajdos, F. F., Yoo, S., Worthylake, D. K., Houseweart, M., Sundquist, W. I. & Hill, C. P. (1996). Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell*, **87**, 1285–1294.
- Wall, M. A., Coleman, D. E., Lee, E., Iniguez-Lluhi, J. A., Posner, B. A., Gilman, A. G. & Sprang, S. R. (1995). The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2. *Cell*, **83**, 1047–1058.
- Neer, E. J., Schmidt, C. J., Nambudripad, R. & Smith, T. F. (1994). The ancient regulatory-protein family of WD-repeat proteins. *Nature*, **371**, 297–300.
- Umen, J. G. & Guthrie, C. (1995). The second catalytic step of pre-mRNA splicing. *RNA*, **1**, 869–885.
- Jiang, J., Horowitz, D. S. & Xu, R. M. (2000). Crystal structure of the functional domain of the splicing factor Prp18. *Proc. Natl Acad. Sci. USA*, **97**, 3022–3027.
- Baker, E. K., Colley, N. J. & Zuker, C. S. (1994). The cyclophilin homolog NinaA functions as a chaperone, forming a stable complex *in vivo* with its protein target rhodopsin. *EMBO J.* **13**, 4886–4895.
- Göthel, S. F. & Marahiel, M. A. (1999). Peptidyl-prolyl *cis-trans* isomerases, a superfamily of ubiquitous folding catalysts. *Cell. Mol. Life*, **55**, 423–436.

26. Wedemeyer, W. J., Welker, E. & Scheraga, H. A. (2002). Proline *cis-trans* isomerization and protein folding. *Biochemistry*, **41**, 14637–14644.
27. Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L. & Lührmann, R. (2002). Small nuclear ribonucleoprotein remodelling during catalytic activation of the spliceosome. *Science*, **298**, 2205–2208.
28. Reuter, K., Nottrott, S., Fabrizio, P., Lührmann, R. & Ficner, R. (1999). Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein. *J. Mol. Biol.* **294**, 515–525.
29. Otwinowski, Z. & Minor, W. (1996). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
30. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect D*, **54**, 905–921.
31. Perrakis, A., Harkiolaki, M., Wilson, K. S. & Lamzin, V. S. (2001). ARP/wARP and molecular replacement. *Acta Crystallog. sect. D*, **57**, 1445–1450.
32. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
33. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
34. Bohm, G., Muhr, R. & Jaenicke, R. (1992). Quantitative analysis of protein far UV circular dichroism spectra by neuronal networks. *Protein Eng.* **5**, 191–195.
35. Berrimann, M. & Fairlamb, A. H. (1998). Detailed characterization of a cyclophilin from the human malaria parasite *Plasmodium falciparum*. *Biochem. J.* **334**, 437–445.
36. Wiseman, T., Williston, S., Brandts, J. F. & Lin, L. N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131–137.
37. Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
38. Barton, G. J. (1993). ALS-CRIP-A tool for multiple sequence alignments. *Protein Eng.* **6**, 37–40.
39. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.

*Edited by R. Huber*

*(Received 21 January 2003; received in revised form 14 May 2003; accepted 21 May 2003)*

*Note added in proof:* After this manuscript had been submitted the orthologues CypH from *Schizosaccharomyces pombe* was identified and characterized by Pemberton *et al.* (*J. Chromatog. B Analyt. Technol. Biomed. Life Sci.* **786**, 81–91).