Published in final edited form as:

Biochem J. ; 474(21): 3689–3704. doi:10.1042/BCJ20170366.

The spliceosomal proteins PPIH and PRPF4 exhibit bi-partite binding

Caroline Rajiv^{1,*}, S. RaElle Jackson^{1,†,*}, Simon Cocklin¹, Elan Z. Eisenmesser², and Tara L. Davis^{1,3}

¹Department of Biochemistry and Molecular Biology, Drexel University College of Medicine, Philadelphia, PA 19102, U.S.A

²Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, U.S.A

³Department of Biochemistry and Molecular Biology, Drexel University College of Medicine, 245 N. 15th St. MS 497, Philadelphia, PA 19125, U.S.A

Abstract

Pre-mRNA splicing is a dynamic, multistep process that is catalyzed by the RNA (ribonucleic acid)–protein complex called the spliceosome. The spliceosome contains a core set of RNAs and proteins that are conserved in all organisms that perform splicing. In higher organisms, peptidyl-prolyl isomerase H (PPIH) directly interacts with the core protein pre-mRNA processing factor 4 (PRPF4) and both integrate into the pre-catalytic spliceosome as part of the tri-snRNP (small nuclear RNA–protein complex) subcomplex. As a first step to understand the protein interactions that dictate PPIH and PRPF4 function, we expressed and purified soluble forms of each protein and formed a complex between them. We found two sites of interaction between PPIH and the N-terminus of PRPF4, an unexpected result. The N-terminus of PRPF4 is an intrinsically disordered region and does not adopt secondary structure in the presence of PPIH. In the absence of an atomic resolution structure, we used mutational analysis to identify point mutations that uncouple these two binding sites and find that mutations in both sites are necessary to break up the complex. A discussion of how this bipartite interaction between PPIH and PRPF4 may modulate spliceosomal function is included.

Correspondence: Tara L. Davis (tara.davis@drexelmed.edu).

*These authors contributed equally to this work.

†Present address: Department of Cancer Biology, Thomas Jefferson University, Philadelphia, PA 19107, U.S.A.

Author Contribution

T.L.D. conceived the project and designed the flow of experiments. C.R. expressed and purified proteins, and performed experiments within Figures 3, 4, 7, and 8, and Supplementary Figures S1 and S4. C.R. and T.L.D. analyzed data from these experiments and jointly created figures. S.R.J. cloned, expressed and purified proteins, and performed experiments within Figures 2, 5, and 6, and Supplementary Figure S2. S.R.J. and T.L.D. analyzed data from these experiments. E.Z.E. directed the NMR experiments described in Supplementary Figure S3 and processed data. S.C. provided access and training for the SPR experiments described in the manuscript. C.R. and T.L.D. wrote the paper with input and editing from all other authors.

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

Introduction

The spliceosome is a complex and dynamic macromolecular machine that removes intronic sequences from messenger RNA (pre-mRNA) and joins coding sequences together. The spliceosome is composed of five uridine-rich small nuclear RNAs (snRNAs) that interact with essential protein chaperones to form snRNP (small nuclear RNA–protein complex) subcomplexes. Human snRNPs contain ~90 conserved spliceosomal proteins, which in addition to chaperones include proteins that assist in recognizing splice sites and positioning the pre-mRNA and snRNAs appropriately for the two catalytic steps of splicing [1–4]. Comparison of purified yeast and metazoan spliceosomes has found additional protein factors that abundantly incorporate into snRNPs and into intermediate and catalytic spliceosomal complexes [5–7]. These additional proteins are likely to play regulatory roles in the organisms they are found in, but many contain no orthologs in lower organisms or homology to annotated domains. This hinders study of their function.

On the other hand, the study of conserved spliceosomal proteins in tractable model organisms has contributed greatly to functional characterization [8,9]. One example is pre-mRNA processing factor 4 (PRPF4), named due to defects in pre-mRNA processing when deleted from *Saccharomyces cerevisiae*. PRPF4 is found as part of the U4/U5/U6 tri-snRNP. Tri-snRNP integrates into the spliceosome by interacting with U2 snRNP, creating the intermediate B complex [4,10]. Mutations in several tri-snRNP factors, including PRPF4, cause autosomal dominant versions of the neurodegenerative disease retinitis pigmentosa (adRP) by compromising tri-snRNP stability [11–14]. The disease-causing mutation is in the highly conserved C-terminal region of PRPF4, which interacts with pre-mRNA processing factor 3 (PRPF3). Destabilization of tri-snRNP results in aberrant mRNA splicing that contributes to the disease. This highlights the contribution of protein–protein interactions to proper spliceosome function, in addition to protein–RNA interactions [11–14].

The N-terminal region of PRPF4 is not conserved in *S. cerevisiae*, but is found in a variety of eukaryotes. A short 30-mer peptide derived from this ~200-residue region was shown to interact with peptidyl-prolyl isomerase H (PPIH), which is also not conserved in *S. cerevisiae* [15–17]. PPIH belongs to a family of proteins called cyclophilins that can catalyze the interconversion between *cis* and *trans* stereoisomers of proline, inducing local changes in protein conformation around the isomerization site [18,19]. Cyclophilins are evolutionarily ancient, with homologs existing in *Escherichia coli*, *Plasmodium falciparum*, and viruses [20–22]. There are eight nuclear-localized cyclophilins in human, and all eight are found abundantly in human spliceosome. PPIH abundantly associates with tri-snRNP and intermediates B splicing complexes [1,4,17,23–25]. Cyclophilins are assumed to function as prolyl isomerases *in vivo*, but previous studies have found no role for the isomerase active site in mediating spliceosomal interactions [16,24,26]. More recently, we showed that incorporation of the nuclear cyclophilins PPIH, PPIL2, CWC27, and PPIG into assembling spliceosome is important for proper spliceosomal assembly and catalytic function *in vitro*. This function was also independent of isomerase activity [27].

A spliceosome-wide siRNA screen recently provided direct evidence that both PRPF4 and PPIH regulate alternative splicing in human cells, and indeed that tri-snRNP proteins are

over-represented in this process [28]. This study indicates that modulating protein–protein interactions within the spliceosome can regulate alternative splicing. As part of a larger goal to understand how spliceosomal proteins interact with each other in order to modulate splicing, we have undertaken an in-depth analysis of the PPIH–PRPF4 interaction using human genes expressed and purified from bacteria. We found a novel interaction between PPIH and PRPF4, and optimized the previous interaction to achieve nanomolar affinity. We characterized these interactions and found a mutational scheme that completely eliminates binding between PPIH and PRPF4. This work identifies a new bipartite binding interface between PPIH and PRPF4, and indicates that there is a role for the isomerase active site in spliceosomal protein interaction.

Experimental procedures

Cloning, expression, and purification of PPIH, PPIA, and PRPF4

Subcloning of full-length PPIH (BC003412) into pET15-MHL vector (EF456738) was described previously [19]. PPIA (BC137057) and PRPF4 (BC001588) constructs were subcloned into pET15-MHL vector using the In-Fusion EcoHD kit (Clontech). Mutant PPIH and PRPF4 constructs were generated using mutagenic PCR primers (see Supplementary Methods for all oligo sequences). After PCR, samples were treated with DpnI (New England Biolabs) for 1 h at 37°C before transforming into DH5α cells (New England Biolabs). Plasmids obtained from positive colonies were purified using the Zymoprep Plasmid Miniprep Kit (Zymo Research) and sequenced to verify mutation.

For protein expression, 50 ng of plasmid DNA was transformed into BL21(DE3) cells (Life Technologies). These cells were grown in Terrific Broth (Research Products International) at 37°C using a LEX bioreactor (Epiphyte), induced using 100 μM IPTG (CellGro) at 15°C, and harvested 12 h later by centrifugation at 4000×g for 20 min. PPIA was prepared in the same manner, except expression was performed at 25°C overnight. Pellets were stored at –80°C until use. Frozen bacterial cell pellets were thawed in warm water and then resuspended in lysis buffer [50 mM Tris (pH 8), 500 mM NaCl, 3 : 1 ratio of BPER : YPER detergents (Thermo Fisher), DNase I, lysozyme, and Protease Inhibitor P2714 (all from Sigma)]. Cells were sonicated (frequency 50% maximum, 30 s on/30 s off, 5 min sonication time, Virsonic).

Immobilized metal ion chromatography (IMAC) purification of PPIH was as in Davis et al. [19], with the substitution of His-Link resin (Promega) for a Ni-NTA resin. The protein contains an N-terminal glycine and serine was added to the native sequence, left behind after protease cleavage of a glutathione-*S*-transferase fusion tag. Original preparations of PRPF4 followed the same purification procedures as those for PPIH, but after identifying PRPF4 as an intrinsically disordered protein, a denaturing purification scheme was optimized in order to remove associated *E. coli* proteins. For this protocol, 8 M urea was added to lysis buffer to make denaturing buffer, and after sonication, lysate was mixed with His-Link resin for 30 min at 4°C. Post-incubation, the resin was placed in a gravity column and washed sequentially with denaturing buffer, wash buffer [denaturing buffer + 10 mM imidazole (pH 8)], and elution buffer [denaturing buffer + 250 mM imidazole (pH 8)]. The eluate was then sequentially dialyzed in dilution buffers 1 [50 mM Tris (pH 8), 500 mM NaCl, 6 M urea, 1

M L-Arginine, and 5 mM β -ME], 2 [50 mM Tris (pH 8), 500 mM NaCl, 4 M urea, 1 M L-Arginine, and 5 mM β -ME], and 3 [50 mM Tris (pH 8), 500 mM NaCl, 2 M urea, 0.05 M L-Arginine, 5% glycerol, and 5 mM β -ME] for 1 h each before being filtered through a 0.45- μ m syringe filter and run over a gel filtration column. Size exclusion chromatography (SEC) for all proteins was accomplished using a Sephadex 200 column (GE Healthcare) on an NGC Chromatography System (Bio-Rad) using gel filtration buffer [50 mM Tris (pH 8), 500 mM NaCl, 5 mM β -mercaptoethanol (β -ME), and 1 mM EDTA]. Fractions were analyzed using SDS-PAGE and pooled. If needed, protein was concentrated using Vivaspin-20 concentrators with a 5000 or 10 000 MWCO cutoff (GE Healthcare). ^{15}N - ^1H heteronuclear single quantum correlation (HSQC) spectroscopy was performed on protein obtained through both soluble and denaturing purification methods, and spectra were identical (data not shown).

Isothermal calorimetry

Experiments were performed in a Nano ITC (TA Instruments). Purified PPIH (40–100 μ M) in isothermal calorimetry (ITC) buffer [50 mM HEPES (pH 8), 200 mM NaCl, and 5 mM β -ME] was placed into the sample cell, and PRPF4 proteins were injected at concentrations of 0.5–1.25 mM. Injections of 1–2 μ l were performed at 25°C with an interval of 1 s and a constant stirring rate of 250 rpm. All samples were degassed for 10–30 min before each run. All proteins were injected into ITC buffer, which was used to perform baseline subtraction. The corrected data were analyzed and the curves were fit to an independent binding model using NanoAnalyze (TA instruments).

Surface plasmon resonance

A ProteOn XPR-36 instrument (Bio-Rad) was used to conduct surface plasmon resonance (SPR) experiments at 25°C. PPIH, W133A PPIH mutant, or PPIA was immobilized on a ProteON GLC chip (Bio-Rad, Hercules, CA). The chip channels were activated by injecting a solution of 25 mM *N*-ethyl-*N'*-(3-dimethylaminopropyl) carbodi-imide (EDC) freshly mixed with 8 mM sulfo-*N*-hydroxysuccinimide (sulfo-NHS) at a flow rate of 50 μ l/min. Ligands were diluted in acetate buffer (pH 5.5, Bio-Rad Laboratories) to a final concentration of 0.5 μ M and immobilized to 2000 response units (RUs). Excess ester groups on the sensor surface were capped with the injection of 1 M ethanolamine-HCl (pH 8.5, Bio-Rad) at a flow rate of 50 μ l/min. Chip equilibration into SPR buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM β -ME, 1 mM EDTA, and 0.01% Tween) was performed overnight. Proteins used as analytes were dialyzed into SPR buffer before use to minimize bulk effects. Binding sensorgrams were collected and initially processed using the integrated ProteOn Manager software (Bio-Rad Laboratories). Analyte responses were corrected for nonspecific interactions using a control human IL-2 (Bio-Rad) immobilized to 2000 RU. Corrected RUs were plotted against analyte protein concentration and fitted to a one-site specific binding model using the GraphPad Prism to obtain equilibrium-binding constants. All results shown are representative of at least three separate experiments.

Circular dichroism

Circular dichroism (CD) spectra of single proteins and complexes were taken using a J-810 Spectropolarimeter (JASCO). Wavelength scans from 196 to 260 nm were carried out using

a 1-mm pathlength cuvette in 30 mM sodium phosphate buffer (pH 7.5) and 75 mM NaCl. A target concentration of 8 μ M was used for single proteins and for complexes after optimization of CD signal and HT voltage values. For complexes, protein concentrations measured after spectra were taken results in predicted binding saturation of 76% (PRPF4₁₋₁₆₉), 90% (PRPF4₁₀₆₋₁₆₉), and 33% (PRPF4₁₋₉₈). Spectra shown are a buffer-subtracted average of three scans and collected over 0.2 nm intervals at 25°C. The web server CAPITO was used to evaluate the % secondary structure and fold state (unfolded, molten globule, and globular) [29,30]. Datasets of folded and unfolded protein CD spectra are available from the CAPITO website for analysis. Thermal denaturation curves were recorded at a wavelength of 220 nm from 25 to 75°C. Three spectra were recorded at every 1°C interval and averaged. Non-linear regression fit to a Boltzmann sigmoidal equation in the GraphPad Prism was used to calculate T_m .

Analytical ultracentrifugation

Sedimentation velocity (SV) and sedimentation equilibrium (SE) experiments were carried out with a ProteomeLab XL-I analytical ultracentrifuge (Beckman Coulter) using a 50 Ti rotor. Concentrations of proteins for SE were 0.3–0.5 mM and for SV from 0.3 to 0.9 mM, depending on the predicted extinction molar coefficient for each protein or complex. Analytical ultracentrifugation analysis (AUC) buffer was 20 mM Tris–HCl (pH 8), 500 mM NaCl, 5 mM β -ME, and 1 mM EDTA. After cell assembly of a six sector centerpiece for SE, 100 μ l of protein was added to the three sample chambers and 120 μ l of buffer was added to each of the three buffer chambers. Samples were spun at 2400, 6600, 15 000, 26 500, 41 500, 60 000, and 80 000 $\times g$ until all samples sedimented (~72 h). Four hundred microliters of protein were added to the SV 2 sector center piece sample chamber and 420 μ l of buffer to the buffer chamber before spinning samples at 135 000 $\times g$ for ~18 h. SV data were fitted using a continuous sedimentation coefficient [$\alpha(s)$] distribution model in SEDFIT [31]. For SE data, molecular mass analysis was performed using the ‘species analysis’ model available in SEDPHAT with RI (radial-invariant) noise baseline correction [32]. Global fit to four to six runs was used to calculate the reported molecular mass. For analysis of complexes, equimolar complexes were formed with a minimum concentration of 300 μ M. This resulted in a minimum predicted binding saturation of 96% (PRPF4₁₋₁₆₉), 98% (PRPF4₁₀₆₋₁₆₉), and 83% (PRPF4₁₋₉₈).

Thermal stability assays

5000 $\times$ SYPRO Orange (Sigma) was diluted to 200 $\times$ in gel filtration buffer [500 mM NaCl, 20 mM Tris (pH 8), 5 mM β -ME, and 1 mM EDTA]. One microliter of dye was added to 9 μ l of purified protein, and a thermal ramp was run from 25 to 99°C in 0.5°C/s increments. Data were collected and analyzed using a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad). Non-linear regression fit to a Boltzmann sigmoidal equation in the GraphPad Prism was used to calculate T_m . For analysis of complexes, PPIH was held at a constant 25 μ M, and PRPF4 proteins were titrated from 25 to 75 μ M. This resulted in a range of predicted binding saturations of 86–99% (PRPF4₁₋₁₆₉), 95–99.8% (PRPF4₁₀₆₋₁₆₉), and 52–83% (PRPF4₁₋₉₈).

Results

Identification of soluble PRPF4 protein for protein–protein interaction analysis with PPIH

PRPF4 is a 522 amino acid protein. The gene is dominated by a C-terminal WD-40 domain (residues 220–522), which includes the site for interaction with PRPF3 and the mutation that causes dominant negative retinitis pigmentosa [12]. The N-terminal region of PRPF4 includes a site of interaction with PPIH [15,17], but initial publications focused on 30 of the N-terminal 220 residues [16] (Figure 1). We undertook expression trials, extending our analysis to as much of the full-length PRPF4 protein as possible. Nine constructs spanning the full-length gene (BC001588) were subcloned into the pET15-MHL vector (Addgene 26092) encoding a cleavable N-terminal hexa-histidine tag, and expression trials were conducted in the BL21(DE3) strain of *E. coli*. All constructs containing the WD-40 domain were insoluble; the longest construct that yielded soluble protein was one encoding for residues 1–169 (referred to throughout as PRPF4_{1–169}). PRPF4_{1–169} was purified to homogeneity using Ni²⁺-IMAC and SEC (Figure 2), with typical yields of 5–10 mg/l of cell culture. Cloning, expression, and purification of the full-length gene for PPIH (BC003412) have been described previously [19].

PRPF4_{1–169} binds to PPIH using two interaction sites

We first addressed the question of whether PRPF4_{1–169} bound to PPIH using ITC. We found that PRPF4_{1–169} bound to PPIH with a K_d value of 0.5 μ M. This is tighter than the 2 μ M affinity reported previously for the PRPF4 peptide consisting of residues 107–136 (Figure 2) [16]. Moreover, the molar ratio indicated that there might be more than one binding site between the two proteins (Figure 2 and Supplementary Table S1). We then purified two more versions of PRPF4 for testing: one encoding the first 98 residues (PRPF4_{1–98}) and one encoding the last 64 residues (PRPF4_{106–169}) (Figure 2). We identified very tight binding between PRPF4_{106–169} and PPIH. The value of K_d of 0.08 μ M is significantly higher affinity than that reported for the PRPF4_{107–136} peptide, and six times tighter than binding to PRPF4_{1–169} (Figure 2). Our ITC trials indicated that PRPF4_{1–98} might also interact with PPIH, but we were unable to reach high enough concentrations of PRPF4_{1–98} to reach saturation (Figure 2). We therefore turned to a complementary technique, SPR.

We immobilized PPIH onto a ProteOn GLC sensor chip, using amine coupling, for use in the Bio-Rad ProteON™ XPR36 system. PRPF4_{1–169}, PRPF4_{106–169}, and PRPF4_{1–98} were used as analytes and flowed over the sensor chip at five to six protein concentrations. We used nonlinear curve fitting to the Langmuir-binding isotherm to obtain K_d (see Methods for details of data analysis). PRPF4_{1–169} interacted with PPIH with an affinity of 0.39 ± 0.04 μ M, and PRPF4_{106–169} with an affinity of 0.08 ± 0.01 μ M (Figure 3). These values are quite similar to those obtained by ITC (Figure 2). PRPF4_{1–98} bound to PPIH with a modest K_d value of 10.84 ± 1.37 μ M (Figure 3). Protein–protein interaction analysis using ITC and SPR revealed a nanomolar affinity-binding site between residues 106–169 of PRPF4 and PPIH. A binding site within the N-terminal 98 residues of PRPF4 was also discovered.

Identification of PRPF4₁₋₁₆₉ as an intrinsically disordered region of PRPF4

The 30-mer peptide from PRPF4 had been previously been crystallized in complex with PPIH (Figure 1) [16]. We tried to crystallize the high-affinity complexes formed between PRPF4₁₀₆₋₁₆₉ and PRPF4₁₋₁₆₉ with PPIH as well. After extensive crystal trials using a variety of complex concentrations and buffer conditions, all crystals obtained in our trials contained density only for PPIH. Both ESpritz [33] and IUPred [34] servers indicated that the entire PRPF4₁₋₁₆₉ region tends toward disorder, which could explain our lack of success in crystallization (Figure 4). To experimentally validate this hypothesis, we used CD to identify α , β , and coil propensity of our PRPF4 proteins. PRPF4₁₋₁₆₉, PRPF4₁₀₆₋₁₆₉, and PRPF4₁₋₉₈ were all found to be predominantly coil as assigned by the CAPITO server [35]. PRPF4₁₋₁₆₉, PRPF4₁₀₆₋₁₆₉, and PRPF4₁₋₉₈ spectra have a classically ‘irregular’ shape, as evidenced by negative ellipticity values in the far UV. There is also a slight propensity for α -helix, especially in PRPF4₁₀₆₋₁₆₉, as shown by the slight deflection at 222 nm (Figure 4). When compared with a database containing CD spectra of folded and disordered proteins, PRPF4₁₋₁₆₉ and PRPF4₁₋₉₈ have ellipticity values at 200 and 222 nm that place them in a region where molten globule proteins cluster (Figure 4) [29,30].

Additional validation of the disordered status of all three PRPF4 proteins was provided by AUC. PPIH, PRPF4₁₋₁₆₉, PRPF4₁₀₆₋₁₆₉, and PRPF4₁₋₉₈ were first characterized using SE-AUC. Since SE-AUC measures the size of the molecule in solution independent of shape, it provides a way to unambiguously define PRPF4 and PPIH as monomers or higher-order oligomers in solution. SE-AUC confirmed that PPIH, PRPF4₁₋₁₆₉, PRPF4₁₀₆₋₁₆₉, and PRPF4₁₋₉₈ were all monomers in solution (Figure 5 and Supplementary Table S3). After establishing that all proteins were monomeric, we examined the shape of our proteins in solution using SV-AUC. SV-AUC provides a measure of the extended nature of a protein in solution using a value called the frictional ratio [31]. Globular proteins are expected to exhibit frictional ratios of 1.2–1.5; numbers higher than this indicate a protein that is extended in solution. PRPF4₁₋₁₆₉ was modeled with an apparent molecular mass of 21.5 kDa, 100.8% of expected, and the frictional coefficient refined to 2.2. PRPF4₁₀₆₋₁₆₉ was found in a monomer : dimer equilibrium, with apparent molecular mass of 9.6 kDa (99.8% of expected) and 22.4 kDa (234%). The modeled frictional coefficient was 2.2. PRPF4₁₋₉₈ was found to have an apparent molecular mass of 18.6 kDa, 140% of the expected molecular mass, and with a modeled frictional ratio of 2.3 (Figure 5 and Supplementary Table S3). We believe that the dimer population seen here for the PRPF4₁₀₆₋₁₆₉ and the larger than expected molecular mass of the PRPF4₁₋₉₈ are largely due to the high concentrations of protein needed to observe signal in SV-AUC, because we do not see any evidence of multimerization in solution using other methods. Additionally, the large values for the frictional ratio are indicative of highly extended shape in solution, as would be expected from intrinsically disordered regions (IDRs), and this may skew the modeling of *s*-values, leading to larger than expected molecular mass.

PRPF4 does not undergo a disorder-to-order transition upon binding PPIH

The 30-mer peptide underwent a disorder-to-order transition upon crystallization with PPIH [16]. Since our three constructs of PRPF4 exhibit a strong tendency toward disorder, we asked whether any, or all, of these regions undergo a disorder-to-order transition upon

interaction with PPIH. We first examined PRPF4–PPIH complexes using CD. We found that the addition of PPIH to PRPF4 results in spectra that are practically identical with that of PPIH alone, suggesting that the addition of PRPF4 to PPIH does not result in appreciable gain of α or β structure (Supplementary Figure S1). Next, we verified using SE-AUC that complexes between PPIH and PRPF4_{1–169} and PRPF4_{1–98} have 1 : 1 stoichiometry. However, we find that there are two molecules of PRPF4_{106–169} in complex with each PPIH. This is likely connected to soluble multimerization of PRPF4_{106–169} at high micromolar concentrations (0.5 mM), as indicated by the SV-AUC data of PRPF4_{106–169} alone (Figure 6 and Supplementary Table S4). For all future work with PRPF4_{106–169}, we only used lower micromolar concentrations, where we have never seen any evidence of PRPF4_{106–169} multimerization. We asked whether formation of complex would change the extended shape of PRPF4 in solution, measured by the frictional ratio. We found for the PRPF4_{106–169} complex, the only one where we could obtain both good signal and fully saturated complex, that there was no change in the frictional ratio compared with PRPF4_{106–169} alone (Supplementary Figure S2 and Table S4). Finally, we performed nuclear magnetic resonance (NMR) experiments on PRPF4_{1–169} alone and in complex with PPIH. ¹⁵N-¹H HSQC spectroscopy exhibited a lack of dispersion along the proton axis, and only a limited number of peaks were visible in the spectra — hallmarks of an unfolded protein or IDR (Supplementary Figure S3). Adding equimolar amount of unlabeled PPIH to the experiment resulted in nearly identical spectra, indicating that PRPF4_{1–169} does not undergo a significant disorder-to-order event when bound to PPIH.

Mutagenesis combined with SPR confirms the interaction surfaces within the PRPF4–PPIH complex, and NMR indicates no isomerization of PRPF4_{1–98}

The data presented above indicate a novel, two-site binding mode for interaction between PRPF4 and PPIH. We next sought to rationally design mutants that might interfere with either or both of these binding sites. Based on the co-complex structure between the 30-mer peptide with PPIH, only one residue was a candidate for point mutation of the high-affinity site: F122, which we changed to alanine (Figure 1). When PRPF4_{1–169} F122A was used as an analyte, there was more than a 10-fold reduction in binding affinity to PPIH (15.6 compared with 0.38 μ M). The binding of PRPF4_{106–169} F122A to wild-type PPIH was also weakened, by ~7-fold (0.63 compared with 0.08 μ M). These results confirm the contribution of F122 to high-affinity interaction, although it is not enough to disrupt either complex entirely (Figure 7). The lack of structural information for the PRPF4_{1–98} region limited our ability to predict effective point mutations to disrupt the low-affinity site. However, if PRPF4_{1–169} remains largely unstructured and extended upon binding PPIH, the contour length would be ~400 Å. If the 30-mer structure approximates the location of the tight binding site on PPIH, then the distance from the most N-terminal residue (E107) to the isomerase active site is ~200 Å, making this a possible secondary binding site. Therefore, we tested a mutation in the isomerase active site of PPIH that lowers binding affinity for proline and abrogates proline isomerization (W133A) [19,36]. We immobilized PPIH W133A and tested wild-type PRPF4_{1–169}, PRPF4_{106–169}, and PRPF4_{1–98}. All three PRPF4 analytes bound to PPIH W133A without loss of affinity (Supplementary Figure S4). We then immobilized PPIH W133A and tested for binding to F122A mutants of PRPF4_{1–169} and PRPF4_{106–169}. PRPF4_{106–169} F122A interacted with PPIH with 1 μ M affinity. We cannot

detect any binding between PPIH W133A and PRPF4_{1–169} F122A (Figure 7). These results indicate that the W133A mutation in PPIH is perturbing an interaction with the PRPF4_{1–98} region, and F122A weakens the interaction with the PRPF4_{106–169} region, but total disruption of the complex is not achieved until both mutations are used.

These findings led us to ask two more questions about the PPIH–PRPF4 interaction. First, the identification of W133A as a binding determinant for PRPF4_{1–98} led us to ask whether this fragment of PRPF4 might be a substrate for proline isomerization by PPIH. There are three potential prolines that might be substrates (P17, P23, and P28). We used an NMR-based assay to check for proline isomerization. In this experiment, called ¹⁵N-¹H ZZ-exchange spectroscopy, the *cis* and *trans* resonances of the residue preceding a proline are monitored; upon the addition of enzyme, additional exchange peaks appear in the spectra due to catalysis and the mixing of magnetization during relaxation [37–40]. As seen in Supplementary Figure S5, we do not see any exchange peaks indicating active isomerization under the standard experimental conditions tested. It is possible that either isomerization is too slow to observe exchange, or that the inherent *cis/trans* isomer population in PRPF4_{1–98} is biased so much toward the *trans* isomer that we cannot see significant exchange between the two forms. This might be expected for prolines contained within a fully unstructured polypeptide structure. For now, we do not currently see isomerization under conditions that have worked for other active cyclophilins and their verified substrates [37–40]. Finally, we tested for specificity between PPIH and PRPF4 fragments by immobilizing PPIA and testing for the ability of PRPF4 to interact with PPIA when compared with binding to PPIH. PPIA is the canonical member of the cyclophilin family and is highly similar to PPIH in amino acid sequence (54% identity and 70% similarity across the 170 residues of the cyclophilin domain). Additionally, all nine of the residues that comprise the isomerase active site are identical between PPIH and PPIA [18]. We used PRPF4_{1–169}, PRPF4_{106–169}, and PRPF4_{106–169} as analytes to test all binding sites for specificity. As seen in Figure 8, we find that all three PRPF4 fragments are exquisitely specific for PPIH, with no observable interaction with PPIA seen.

Formation of the PPIH–PRPF4 complex significantly increases thermal stability

We have described a tight binding site between PPIH and PRPF4_{106–169}, and a weak binding site in PRPF4_{1–98}. We wondered how these interactions were contributing to the overall stability of the complex. We used the ThermoFluor technique, utilizing SYPRO Orange binding to exposed hydrophobic regions upon heating, to assess the effect of PRPF4 binding on PPIH stability [41]. We tested PRPF4_{1–169}, PRPF4_{106–169}, PRPF4_{1–98}, and PPIH proteins alone. In this assay, we found no significant melting transition for PRPF4_{1–169}, PRPF4_{106–169}, or PRPF4_{1–98} (Figure 9). PPIH exhibited a cooperative transition with an average melting temperature of 53.8°C. We then performed a series of titrations with PRPF4_{1–169}, PRPF4_{106–169}, or PRPF4_{1–98}, using protein concentrations to give 80–99% saturation (described in Methods). We found that binding to PRPF4_{1–169} resulted in an upward shift of almost 12°, to an average of 65.5°C. This highly stabilizing effect was predominantly due to PRPF4_{106–169}, which induced stabilization up to an average of 63.8°C. Binding to PRPF4_{1–98} did not impart any additional thermal stability to PPIH (Figure 9). We verified these findings using CD, and obtained practically identical results, with a thermal

stability shift of 6°C (Figure 9). We conclude that binding to PRPF4_{1–169}, and especially to the high-affinity-binding site in PRPF4_{106–169}, creates a highly stabilized PPIH–PRPF4 complex.

Discussion

PRPF4_{1–169} binds to PPIH using two unique, conserved interaction sites

We have discovered that the interaction between PPIH and the N-terminal region of PRPF4 is more complicated than previously reported. Our work identifies a nanomolar affinity-binding site contained within PRPF4_{106–169}. We have also identified a micromolar affinity site between PPIH and PRPF4_{1–98}. Conservation of both these sites in PRPF4 genes indicates that both are relevant to function. The C-terminal WD-40 domain of PRPF4 is found in all organisms that retain the ability to splice mRNA, including *S. cerevisiae*. However, the yeast PRPF4 gene does not include the N-terminal 200 residues currently under study. *S. cerevisiae* also does not encode for a tri-snRNP-associated ortholog of PPIH. We therefore performed sequence alignment on selected eukaryotes that encode both PRPF4 and a paralog of PPIH. This group included animals, fungi, ameba, green algae, plants, heterokonts, and an alveolate (Supplementary Figure S6) [42,43]. The region from residues 90–190 (human PRPF4 numbering) is broadly conserved throughout eukaryotes (Supplementary Figure S6). The first 100 residues of PRPF4 are less conserved but are present in insects, vertebrates, and at least one annelid. This level of conservation suggests that there is a biological role for the first 100 residues of PRPF4.

The strong effect of mutating F122 indicates that this residue in PRPF4 is critical for high-affinity interaction with PPIH. However, we could not find a point mutation in PPIH with an equally strong effect, although since PRPF4_{106–169} is specific for PPIH there must be some selective determinant (Figure 8). One report found that binding between full-length *in vitro* translated PRPF4 and PPIH, mapped using loop swapping, depended on a region proximal to the one visualized in the 30-mer crystal structure [44]. It is tempting to speculate that the longer polypeptide encoded by PRPF4_{106–169} might interact with this region of PPIH (α 1– β 3 region), explaining the increase in affinity for PPIH we see experimentally. This would be unlikely if the α -helical conformation of the 30-mer crystal was accurate, as the two short helical sections fold back upon one another (Figure 1), pointing away from the α 1– β 3 region. Our data would also support this report, since the α 1– β 3 region is not conserved between PPIH and PPIA. Future work should address this question by further mutational analysis of PPIH to confirm the role of the α 1– β 3 region in PRPF4 interaction.

An active site mutation in PPIH, W133, also alters binding to PRPF4. This is likely due to interaction between a proline residue in PRPF4 and the active site of PPIH [19,36]. There are five prolines in PRPF4_{1–169}, three of which are found in the first 30 residues. Although we cannot formally exclude the prolines at positions 118 and 126, they are present in the PRPF4_{106–169} fragment, which binds with nearly equal affinity to wild-type and W133A forms of PPIH. We, therefore, believe that it is likely the prolines within the first 30 residues of PRPF4 that represent part of the weak-affinity-binding site for PPIH. This is the first direct interaction found between a spliceosomal protein and the active site of a spliceosomal cyclophilin, although we note that there are multiple examples of isomerase-dependent

interactions with other nuclear proteins [45–49]. PRPF4_{1–98} is specific for PPIH over PPIA, indicating that there are interactions outside the isomerase active site that are important for this interaction (Figure 8). On the other hand, we see no evidence that catalyzed isomerization is occurring in the PRPF4_{1–98} fragment (Supplementary Figure S5). Therefore, our model focuses on the binding of prolines, but not catalysis.

PRPF4_{1–169} is an IDR of PRPF4 and does not undergo a disorder-to-order transition upon binding PPIH

The 30-mer peptide co-crystallized with PPIH and adopted an α -helical conformation while interacting with a surface on PPIH opposite to the active isomerase site [16]. Although we identified a much higher affinity-binding site within PRPF4_{106–169}, this extended region does not adopt local conformation in solution, nor upon binding to PPIH. In fact, the entire N-terminal region of PRPF4 is intrinsically disordered and does not order upon binding PPIH. Perhaps, this explains why recent tri-snRNP structures show no ordered density for the first 200 residues of PRPF4 [50].

Model and relevance of the PPIH–PRPF4_{1–169} interaction

Taken altogether, we arrive at a model for the PRPF4–PPIH complex in which an IDR of PRPF4 encodes two interactions with PPIH: a low-affinity site localized in the first 100 residues and a high-affinity site localized in residues 100–200. Why would both these interactions exist and be conserved across eukaryotes? We believe that the two sites serve unique functions in the context of the spliceosomal machinery. First, disease-causing mutations in PRPF genes disrupt protein–protein interactions within the tri-snRNP. Weakening these interactions leads to aberrant splicing and defects in spliceosomal assembly, and then disease [11–14]. Our previous work indicates that PPIH also is needed for proper functioning of human spliceosomal complexes, and a recent siRNA study found that PPIH knockdown results in a host of regulated splicing changes [27,28]. Therefore, we believe that the high-affinity site in PRPF4_{106–169} anchors PPIH into tri-snRNP during assembly and integration into B complex, assuring proper spliceosome function. The anchoring site should be outside of the isomerase active site, as this would prevent transient protein–protein interactions, which are likely mediated by PPIH, from dissociating PPIH out of spliceosomal complexes. Whatever function PPIH plays within the tri-snRNP and B complex, this tethering interaction is very highly conserved in eukaryotes, indicating its importance (Supplementary Figure S6).

The conservation pattern of the N-terminal 100 residues of PRPF4 indicates that it also serves an important biological role, but the larger question is why this region would bind so weakly and yet specifically to PPIH. The addition of a second binding site which lessens the overall affinity of PRPF4 for PPIH implies some sort of auto-inhibition model. Perhaps, in the resting state, PRPF4_{1–98} is weakly bound to PPIH through interactions with the isomerase active site. This would prevent the nonspecific binding of other proline residues by PPIH. This is supported by the finding that the prolines within PRPF4_{1–98} are not being turned over by PPIH, even though the overall affinity of PRPF4_{1–98} for PPIH is similar, if not higher, than that seen for most cyclophilin substrates [51–53]. Binding without catalysis might prevent release of product, which might block the autoinhibitory effect

(Supplementary Figure S5). However, there are likely other residues within PRPF4₁₋₉₈ that participate in binding, as the W133 mutation of PPIH does not abrogate PRPF4₁₋₉₈ binding, while substitution of PPIA which is identical in the proline-binding pocket does. IDRs with repetitive sequences often make a large number of low-affinity interactions with their binding partners, which is a way to maintain affinity through binding to multiple, interchangeable residues in a short region of sequence [54,55]. These IDRs, which are often highly extended, unstructured, and have repetitive polar/charged polypeptide sequences, are considered good targets for post-translational modification (PTM) [56]. PRPF4₁₋₁₆₉ is highly enriched in arginine, lysine, serine, threonine, and tyrosine residues (34% of all residues). NetPhos predicts high-confidence sites for serine/threonine and tyrosine phosphorylation throughout the entire region [57]. By comparison, PhosphoSite Plus, a database of experimentally observed PTMs, indicates five acetylation and SUMOylation sites, and three sites of phosphorylation, in PRPF4₁₋₉₈ [58,59]. However, only two acetylation sites and one phosphorylation site has been observed in PRPF4₁₀₆₋₁₆₉, although the number of target residues is roughly the same. One explanation for this observation is that there is relatively limited accessibility of PTM sites in the high-affinity region while bound to PPIH and the tri-snRNP. The role of the poor-affinity site could then allow PRPF4₁₋₉₈ to be regulated by PTMs, which would then cause it to release partially or completely from PPIH in order to bind some other regulatory element or component of the spliceosome. Multiple PTMs might even be necessary to overcome the promiscuity of binding, but regardless PPIH and PRPF4 would remain tethered to the spliceosome through the high-affinity site. This release of PRPF4₁₋₉₈ from PPIH would then allow other prolinecontaining substrates to interact with PPIH (Figure 10). Verification of this model will require the development of assays in cells to measure the effect of mutations and PTMs on splicing function.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This project was supported by funding from the National Institutes of General Medical Sciences [R00GM094293] (S.R.J. and T.L.D.). E.Z.E. is supported by funding from the National Institutes of General Medical Sciences [R01GM107262]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. This work includes experiments carried out at the Sidney Kimmel Cancer Center X-ray Crystallography and Molecular Characterization shared resource at Thomas Jefferson University, which is supported in part by NCI Cancer Center Support Grant P30 CA56036-17.

Dr Anshul Bhardwaj of TJU provided critical training and expert advice on the ITC, AUC, and CD experiments described in the manuscript, along with helpful comments. NMR experiments were performed at the Rocky Mountain Regional NMR Facility at the University of Colorado, under the supervision of Geoffrey Armstrong, who provided helpful advice on sample preparation and experimental design. The PPIH construct was created while T.L.D. was a member of the Structural Genomics Consortium (SGC, Toronto, Ontario, Canada; <http://thesgc.com>).

Abbreviations

AUC analytical ultracentrifugation analysis

β-ME	beta-mercaptoethanol
CD	circular dichroism
HSQC	¹⁵ N- ¹ H heteronuclear single quantum correlation
IDR	intrinsically disordered region
IMAC	Immobilized metal ion chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal calorimetry
mRNA	messenger RNA
NMR	nuclear magnetic resonance
PPIH	peptidyl-prolyl isomerase H
PRPF3	pre-mRNA processing factor 3
PRPF4	pre-mRNA processing factor 4
PTM	post-translational modification
RNA	ribonucleic acid
RU_s	response units
SE	sedimentation equilibrium
SEC	size-exclusion chromatography
SPR	surface plasmon resonance
snRNA	small nuclear RNA
snRNP	small nuclear RNA–protein complex
SV	sedimentation velocity

References

1. Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell*. 2009; 136:701–718. <https://doi.org/10.1016/j.cell.2009.02.009>. [PubMed: 19239890]
2. Will CL, Lührmann R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 2011; 3
3. Pleiss JA, Whitworth GB, Bergkessel M, Guthrie C. Transcript specificity in yeast pre-mRNA splicing revealed by mutations in core spliceosomal components. *PLoS Biol.* 2007; 5:e90. <https://doi.org/10.1371/journal.pbio.0050090>. [PubMed: 17388687]
4. Agafonov DE, Deckert J, Wolf E, Odenwalder P, Bessonov S, Will CL, et al. Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol. Cell. Biol.* 2011; 31:2667–2682. <https://doi.org/10.1128/MCB.05266-11>. [PubMed: 21536652]

5. Chen Y-IG, Moore RE, Ge HY, Young MK, Lee TD, Stevens SW. Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res.* 2007; 35:3928–3944. <https://doi.org/10.1093/nar/gkm347>. [PubMed: 17537823]
6. Fabrizio P, Dannenberg J, Dube P, Kastner B, Stark H, Urlaub H, et al. The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell.* 2009; 36:593–608. <https://doi.org/10.1016/j.molcel.2009.09.040>. [PubMed: 19941820]
7. Herold N, Will CL, Wolf E, Kastner B, Urlaub H, Luhrmann R. Conservation of the protein composition and electron microscopy structure of *Drosophila melanogaster* and human spliceosomal complexes. *Mol. Cell. Biol.* 2009; 29:281–301. <https://doi.org/10.1128/MCB.01415-08>. [PubMed: 18981222]
8. Bjørn SP, Soltyk A, Beggs JD, Friesen JD. PRP4 (RNA4) from *Saccharomyces cerevisiae*: its gene product is associated with the U4/U6 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.* 1989; 9:3698–3709. <https://doi.org/10.1128/MCB.9.9.3698>. [PubMed: 2528686]
9. Hartwell LH. Macromolecule synthesis in temperature-sensitive mutants of yeast. *J. Bacteriol.* 1967; 93:1662–1670. [PubMed: 5337848]
10. Deckert J, Hartmuth K, Boehringer D, Behzadnia N, Will CL, Kastner B, et al. Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.* 2006; 26:5528–5543. <https://doi.org/10.1128/MCB.00582-06>. [PubMed: 16809785]
11. Chen X, Liu Y, Sheng X, Tam POS, Zhao K, Chen X, et al. PRPF4 mutations cause autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* 2014; 23:2926–2939. <https://doi.org/10.1093/hmg/ddu005>. [PubMed: 24419317]
12. Linder B, Hirmer A, Gal A, Rüther K, Bolz HJ, Winkler C, et al. Identification of a PRPF4 loss-of-function variant that abrogates U4/U6.U5 tri-snRNP integration and is associated with retinitis pigmentosa. *PLoS ONE.* 2014; 9:e111754. <https://doi.org/10.1371/journal.pone.0111754>. [PubMed: 25383878]
13. Mordes D, Luo X, Kar A, Kuo D, Xu L, Fushimi K, et al. Pre-mRNA splicing and retinitis pigmentosa. *Mol. Vis.* 2006; 12:1259–1259. [PubMed: 17110909]
14. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev.* 2003; 17:419–437. <https://doi.org/10.1101/gad.1048803>. [PubMed: 12600935]
15. Horowitz DS, Kobayashi R, Krainer AR. A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *RNA.* 1997; 3:1374–1387. [PubMed: 9404889]
16. Reidt U, Wahl MC, Fasshauer D, Horowitz DS, Luhrmann R, Ficner R. Crystal structure of a complex between human spliceosomal cyclophilin H and a U4/U6 snRNP-60K peptide. *J. Mol. Biol.* 2003; 331:45–56. [https://doi.org/10.1016/S0022-2836\(03\)00684-3](https://doi.org/10.1016/S0022-2836(03)00684-3). [PubMed: 12875835]
17. Teigelkamp S, Achsel T, Mundt C, Göthel SF, Cronshagen U, Lane WS, et al. The 20kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *RNA.* 1998; 4:127–141. [PubMed: 9570313]
18. Daum S, Schumann M, Mathea S, Aumüller T, Balsley MA, Constant SL, et al. Isoform-specific inhibition of cyclophilins. *Biochemistry.* 2009; 48:6268–6277. <https://doi.org/10.1021/bi9007287>. [PubMed: 19480458]
19. Davis TL, Walker JR, Campagna-Slater V, Finerty PJ, Paramanathan R, Bernstein G, et al. Structural and biochemical characterization of the human cyclophilin family of peptidyl-prolyl isomerases. *PLoS Biol.* 2010; 8:e1000439. <https://doi.org/10.1371/journal.pbio.1000439>. [PubMed: 20676357]
20. Galat A, Bua J. Molecular aspects of cyclophilins mediating therapeutic actions of their ligands. *Cell. Mol. Life Sci.* 2010; 67:3467–3488. <https://doi.org/10.1007/s00018-010-0437-0>. [PubMed: 20602248]
21. Kumari S, Roy S, Singh P, Singla-Pareek SL, Pareek A. Cyclophilins: proteins in search of function. *Plant Signal. Behav.* 2013; 8:e22734. <https://doi.org/10.4161/psb.22734>. [PubMed: 23123451]

22. Thai V, Renesto P, Fowler CA, Brown DJ, Davis T, Gu W, et al. Structural, biochemical, and *in vivo* characterization of the first virally encoded cyclophilin from the Mimivirus. *J. Mol. Biol.* 2008; 378:71–86. <https://doi.org/10.1016/j.jmb.2007.08.051>. [PubMed: 18342330]
23. Jurica MS. Detailed close-ups and the big picture of spliceosomes. *Curr. Opin. Struct. Biol.* 2008; 18:315–320. <https://doi.org/10.1016/j.sbi.2008.05.005>. [PubMed: 18550358]
24. Stegmann CM, Lührmann R, Wahl MC. The crystal structure of PPIL1 bound to cyclosporine A suggests a binding mode for a linear epitope of the SKIP protein. *PLoS ONE.* 2010; 5:e10013. <https://doi.org/10.1371/journal.pone.0010013>. [PubMed: 20368803]
25. Cvitkovic I, Jurica MS. Spliceosome database: a tool for tracking components of the spliceosome. *Nucleic Acids Res.* 2013; 41:D132–D141. <https://doi.org/10.1093/nar/gks999>. [PubMed: 23118483]
26. Wang X, Zhang S, Zhang J, Huang X, Xu C, Wang W, et al. A large intrinsically disordered region in SKIP and its disorder-order transition induced by PPIL1 binding revealed by NMR. *J. Biol. Chem.* 2010; 285:4951–4963. <https://doi.org/10.1074/jbc.M109.087528>. [PubMed: 20007319]
27. Adams BM, Coates MN, Jackson SR, Jurica MS, Davis TL. Nuclear cyclophilins affect spliceosome assembly and function *in vitro*. *Biochem. J.* 2015; 469:223–233. <https://doi.org/10.1042/BJ20150396>. [PubMed: 25967372]
28. Papasaikas P, Tejedor JR, Vigevari L, Valcárcel J. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol. Cell.* 2015; 57:7–22. <https://doi.org/10.1016/j.molcel.2014.10.030>. [PubMed: 25482510]
29. Whitmore L, Miles AJ, Mavridis L, Janes RW, Wallace BA. PCDDb: new developments at the protein circular dichroism data bank. *Nucleic Acids Res.* 2017; 45:D303–D307. <https://doi.org/10.1093/nar/gkw796>. [PubMed: 27613420]
30. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 2002; 11:739–756. <https://doi.org/10.1110/ps.4210102>. [PubMed: 11910019]
31. Dam J, Velikovskiy CA, Mariuzza RA, Urbanke C, Schuck P. Sedimentation velocity analysis of heterogeneous protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions *c(s)*. *Biophys. J.* 2005; 89:619–634. <https://doi.org/10.1529/biophysj.105.059568>.
32. Schuck P, Rossmanith P. Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers.* 2000; 54:328–341. [https://doi.org/10.1002/1097-0282\(20001015\)54:5<328::AID-BIP40>3.0.CO;2-P](https://doi.org/10.1002/1097-0282(20001015)54:5<328::AID-BIP40>3.0.CO;2-P). [PubMed: 10935973]
33. Walsh I, Martin AJM, Di Domenico T, Tosatto SCE. ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics.* 2012; 28:503–509. <https://doi.org/10.1093/bioinformatics/btr682>. [PubMed: 22190692]
34. Dosztanyi Z, Csizmek V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.* 2005; 21:3433–3434. <https://doi.org/10.1093/bioinformatics/bti541>. [PubMed: 15955779]
35. Wiedemann C, Bellstedt P, Görlach M. CAPITO—a web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics.* 2013; 29:1750–1757. <https://doi.org/10.1093/bioinformatics/btt278>. [PubMed: 23681122]
36. Bossard MJ, Koser PL, Brandt M, Bergsma DJ, Levy MA. A single Trp121 to Ala121 mutation in human cyclophilin alters cyclosporin A affinity and peptidyl-prolyl isomerase activity. *Biochem. Biophys. Res. Commun.* 1991; 176:1142–1148. [https://doi.org/10.1016/0006-291X\(91\)90404-U](https://doi.org/10.1016/0006-291X(91)90404-U). [PubMed: 2039499]
37. Schlegel J, Redzic JS, Porter CC, Yurchenko V, Bukrinsky M, Labeikovskiy W, et al. Solution characterization of the extracellular region of CD147 and its interaction with its enzyme ligand cyclophilin A. *J. Mol. Biol.* 2009; 391:518–535. <https://doi.org/10.1016/j.jmb.2009.05.080>. [PubMed: 19500591]
38. Camilloni C, Sahakyan AB, Holliday MJ, Isern NG, Zhang F, Eisenmesser EZ, et al. Cyclophilin A catalyzes proline isomerization by an electrostatic handle mechanism. *Proc. Natl Acad. Sci. U.S.A.* 2014; 111:10203–10208. <https://doi.org/10.1073/pnas.1404220111>. [PubMed: 24982184]
39. Doshi U, Holliday MJ, Eisenmesser EZ, Hamelberg D. Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc. Natl Acad.*

- Sci. U.S.A. 2016; 113:4735–4740. <https://doi.org/10.1073/pnas.1523573113>. [PubMed: 27071107]
40. Holliday MJ, Camilloni C, Armstrong GS, Vendruscolo M, Eisenmesser EZ. Networks of dynamic allostery regulate enzyme function. *Structure*. 2017; 25:276–286. <https://doi.org/10.1016/j.str.2016.12.003>. [PubMed: 28089447]
 41. Nettleship JE, Brown J, Groves MR, Geerlof A. Methods for protein characterization by mass spectrometry, thermal shift (ThermoFluor) assay, and multiangle or static light scattering. *Methods Mol. Biol.* 2008; 426:299–318. https://doi.org/10.1007/978-1-60327-058-8_19. [PubMed: 18542872]
 42. Csuros M, Rogozin IB, Koonin EV, Ponting CP. A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS Comput. Biol.* 2011; 7:e1002150. <https://doi.org/10.1371/journal.pcbi.1002150>. [PubMed: 21935348]
 43. Rogozin IB, Carmel L, Csuros M, Koonin EV. Origin and evolution of spliceosomal introns. *Biol. Direct*. 2012; 7:11. <https://doi.org/10.1186/1745-6150-7-11>. [PubMed: 22507701]
 44. Ingelfinger D, Gothel SF, Marahiel MA, Reidt U, Ficner R, Luhrmann R, et al. Two protein-protein interaction sites on the spliceosome-associated human cyclophilin CypH. *Nucleic Acids Res.* 2003; 31:4791–4796. <https://doi.org/10.1093/nar/gkg660>. [PubMed: 12907720]
 45. Hom RA, Chang P-Y, Roy S, Musselman CA, Glass KC, Selezneva AI, et al. Molecular mechanism of MLL PHD3 and RNA recognition by the Cyp33 RRM domain. *J. Mol. Biol.* 2010; 400:145–154. <https://doi.org/10.1016/j.jmb.2010.04.067>. [PubMed: 20460131]
 46. Park S, Osmers U, Raman G, Schwantes RH, Diaz MO, Bushweller JH. The PHD3 domain of MLL acts as a CYP33-regulated switch between MLL-mediated activation and repression. *Biochemistry*. 2010; 49:6576–6586. <https://doi.org/10.1021/bi1009387>. [PubMed: 20677832]
 47. Wang Z, Song J, Milne TA, Wang GG, Li H, Allis CD, et al. Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to Cyp33 and HDAC-mediated repression. *Cell*. 2010; 141:1183–1194. <https://doi.org/10.1016/j.cell.2010.05.016>. [PubMed: 20541251]
 48. Kuraoka I, Ito S, Wada T, Hayashida M, Lee L, Saijo M, et al. Isolation of XAB2 complex involved in pre-mRNA splicing, transcription, and transcription-coupled repair. *J. Biol. Chem.* 2008; 283:940–950. <https://doi.org/10.1074/jbc.M706647200>. [PubMed: 17981804]
 49. Gustafson CL, Parsley NC, Asimgil H, Lee H-W, Ahlback C, Michael AK, et al. A slow conformational switch in the BMAL1 transactivation domain modulates circadian rhythms. *Mol. Cell*. 2017; 66:447–457.e7. <https://doi.org/10.1016/j.molcel.2017.04.011>. [PubMed: 28506462]
 50. Agafonov DE, Kastner B, Dybkov O, Hofele RV, Liu W-T, Urlaub H, et al. Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science*. 2016; 351:1416–1420. <https://doi.org/10.1126/science.aad2085>. [PubMed: 26912367]
 51. Schmid FX. Prolyl isomerase: enzymatic catalysis of slow protein-folding reactions. *Annu. Rev. Biophys. Biomol. Struct.* 1993; 22:123–143. <https://doi.org/10.1146/annurev.bb.22.060193.001011>. [PubMed: 7688608]
 52. Coelmont L, Hanouille X, Chatterji U, Berger C, Snoeck J, Bobardt M, et al. DEB025 (Alisporivir) inhibits hepatitis C virus replication by preventing a cyclophilin A induced *cis-trans* isomerisation in domain II of NS5A. *PLoS ONE*. 2010; 5:e13687. <https://doi.org/10.1371/journal.pone.0013687>. [PubMed: 21060866]
 53. Bosco DA, Eisenmesser EZ, Clarkson MW, Wolf-Watz M, Labeikovsky W, Millet O, et al. Dissecting the microscopic steps of the cyclophilin A enzymatic cycle on the biological HIV-1 capsid substrate by NMR. *J. Mol. Biol.* 2010; 403:723–738. <https://doi.org/10.1016/j.jmb.2010.08.001>. [PubMed: 20708627]
 54. Tompa P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* 2012; 37:509–516. <https://doi.org/10.1016/j.tibs.2012.08.004>. [PubMed: 22989858]
 55. Uversky VN. The multifaceted roles of intrinsic disorder in protein complexes. *FEBS Lett.* 2015; 589:2498–2506. <https://doi.org/10.1016/j.febslet.2015.06.004>. [PubMed: 26073257]
 56. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 2014; 114:6589–6631. <https://doi.org/10.1021/cr400525m>. [PubMed: 24773235]

57. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 1999; 294:1351–1362. <https://doi.org/10.1006/jmbi.1999.3310>. [PubMed: 10600390]
58. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. Phosphositeplus, 2014; mutations, PTMs and recalibrations. *Nucleic Acids Res.* 2015; 43:D512–D520. <https://doi.org/10.1093/nar/gku1267>. [PubMed: 25514926]
59. Matlock MK, Holehouse AS, Naegle KM. Proteomescout: a repository and analysis resource for post-translational modifications and proteins. *Nucleic Acids Res.* 2015; 43:D521–D530. <https://doi.org/10.1093/nar/gku1154>. [PubMed: 25414335]

Figure 1. Identification of soluble PRPF4 proteins for complex with PPIH

(**A**) A 30-mer peptide from PRPF4 interacts with PPIH without contacting the PPIH active site. (Left) A cartoon representation of PPIH (teal) shown interacting with a 30-residue peptide isolated from PRPF4 (goldenrod, residues 106–137) (PDB:1MZW) [16]. The isomerase active surface of PPIH is shown in stick representation, as is the PRPF4 residue F122 mentioned in the text. (Middle) The model is rotated to emphasize the distance between F122 within PRPF4 and the PPIH residue W133. (Right) PPIH is shown in electrostatic surface representation. F122 from PRPF4 inserts into a pocket formed in part by the PPIH $\alpha 1$ – $\beta 3$ loop. (**B**) Domain architecture of PRPF4. The splicing factor motif (SFM), which includes residues 103–155, is defined by the SMART domain server (<http://smart.embl-heidelberg.de/>). This motif includes the sequence of the 30-mer peptide shown in (**A**).

Figure 2. ITC indicates bipartite binding between PRPF4 and PPIH proteins

(A) Graphical representation of constructs used in the current study. (B and C) SEC was performed on (B) PPIH (teal), (C) PRPF4₁₋₁₆₉ (fuchsia), PRPF4₁₀₆₋₁₆₉ (sunset), and PRPF4₁₋₉₈ (blue). Representative chromatograms are shown. All four proteins are well-behaved and homogeneous, although PRPF4₁₋₁₆₉ and PRPF4₁₋₉₈ exhibit retention volumes indicative of extended protein. (C, middle) Representative gels showing SEC fractions of purified proteins. (C, bottom) ITC indicates that PRPF4₁₋₁₆₉ interacts with PPIH with 0.48 μM affinity and two potential binding sites. A K_d value of 0.08 μM is modeled for PRPF4₁₀₆₋₁₆₉ interaction with PPIH, and affinity was not calculated for PRPF4₁₋₉₈ as saturation was not reached.

Figure 3. SPR confirms bipartite binding between PRPF4 and PPIH

SPR was used to validate interactions between (A) PRPF4₁₋₁₆₉, (B) PRPF4₁₀₆₋₁₆₉, and (C) PRPF4₁₋₉₈ with PPIH. (Top) Normalized sensorgrams using PPIH as a ligand and PRPF4 proteins, as indicated, as analytes. (Bottom) Data fit to the Langmuir one-site model to obtain equilibrium-binding curves and calculation of K_d . At least five concentrations were tested to obtain the RUs per concentration reported. Error bars are standard deviations of at least three separate runs.

Figure 4. The PPIH-interacting regions of PRPF4 are intrinsically disordered

(A) Prediction of disorder using the IUPRED server [34]. Residues that fall below the dotted line tend toward order, as exemplified by the trace for PPIH. PRPF4₁₋₁₆₉ has a high tendency toward disorder for residues 1–20 and 150–169, and also tends generally toward disorder throughout. (B) CD spectra of PRPF4₁₋₁₆₉, PRPF4₁₀₆₋₁₆₉, and PRPF4₁₋₉₈ are indicative of minimal or no structured elements. PPIH, a mostly β -protein, is shown for comparison. (C) Comparison of molar ellipticity at 222 and 200 nm places PRPF4 in the ‘molten globule’ region. Reference CD spectra are from refs [29,30]. PRPF4₁₀₆₋₁₆₉ is not included in this analysis due to accuracy issues in measuring the protein concentration of the sample used in analysis.

Figure 5. The PPIH-interacting regions of PRPF4 are monomeric and extended in solution (A–D) SE-AUC data for PRPF4 proteins and for PPIH. All four proteins are modeled as monomers in solution. (E–H) SV-AUC for PRPF4 proteins and for PPIH. PRPF4₁₀₆₋₁₆₉ (F) shows an apparent monomer : dimer equilibrium under the relatively high concentrations of protein used in this experiment. All other proteins are monomeric. The frictional ratio for all PRPF4 proteins are higher than 2, indicating highly extended conformations in solution.

Figure 6. PRPF4 complexes with PPIH largely have 1 : 1 stoichiometry
(A–C) SE-AUC data for PRPF4 complexes. PRPF4_{106–169} (B) complex is modeled as 2 : 1 stoichiometry, in concordance with the data shown in Figure 5.

Figure 7. Mutational analysis of the PRPF4–PPIH complex

SPR was used to interrogate interactions between mutant PRPF4 (F122A) and either wild-type or mutant PPIH. (Top) Normalized sensorgrams using PPIH as a ligand and PRPF4 as analytes. (Bottom) Fitting sensorgram data to the Langmuir one-site model in order to obtain equilibrium-binding curves and calculation of K_d . In all cases, at least five concentrations of PRPF4 proteins were tested in at least three runs to obtain the RUs per concentration reported in the bottom panels. **(A)** PRPF4₁₀₆₋₁₆₉ F122A interacts with wild-type PPIH. **(B)** PRPF4₁₀₆₋₁₆₉ F122A interacts with PPIH mutant W133A. **(C)** PRPF4₁₋₁₆₉ F122A interacts with wild-type PPIH. **(D)** PRPF4₁₋₁₆₉ F122A does not interact with PPIH mutant W133A.

Figure 8. Specificity of the PPIH-PRPF4 interaction

SPR was used to interrogate interactions between PPIH and the canonical cyclophilin PPIA for each of the PRPF4 fragments. (Left) Normalized sensorgrams using PPIH as a ligand and PRPF4 fragments as analytes. (Right) Normalized sensorgrams using PPIA as a ligand and PRPF4 fragments as analytes. (A) PPIH binds to PRPF4₁₋₁₆₉, while PPIA binding is undetectable; (B) PPIH binds to PRPF4₁₀₆₋₁₆₉, while PPIA binding is undetectable; (C) PPIH binds to PRPF4₁₋₉₈, while PPIA binding is undetectable. Similar ranges of analyte concentrations were used for PRPF4₁₋₁₆₉, PRPF4₁₋₉₈, and PRPF4₁₀₆₋₁₆₉ as in Figures 3 and 7.

Figure 9. PRPF4 interaction with PPIH increases thermal stability of the complex

(A and B) Thermal stability assays using SYPRO Orange, which interacts with hydrophobic residues. PPIH protein alone exhibits a cooperative unfolding transition at 54°C, while PRPF4 proteins show no transition (traces not shown). Complexes of either PRPF4₁₋₁₆₉ or PRPF4₁₀₆₋₁₆₉ with PPIH result in a highly stabilized complex with T_m of 66 and 64°C, respectively. (C and D) CD performed under thermal denaturing conditions. In both panels, buffer-subtracted spectra for PPIH and the PRPF4 protein are shown, along with the corresponding complex. Helicity as approximated by molar residue ellipticity at 222 nm is monitored throughout. The effect of interaction with PRPF4₁₋₁₆₉ or PRPF4₁₀₆₋₁₆₉ is to stabilize PPIH by 6°C. No stabilization effect was seen for PRPF4₁₋₉₈ complexes. Note, in (D), a small amount of helical signal in PRPF4₁₀₆₋₁₆₉ protein, but not in PRPF4₁₋₁₆₉, that is destabilized by increasing temperature.

Figure 10. Model for the role of bipartite binding in the PRPF4-PPIH complex
See the section 'Discussion' for details.