



**HAL**  
open science

## Optimizing the First TPR Domain of the Human SPAG1 Protein Provides Insight into the HSP70 and HSP90 Binding Properties

Sana Dermouche, Marie-Eve Chagot, Xavier Manival, Marc Quinternet

► **To cite this version:**

Sana Dermouche, Marie-Eve Chagot, Xavier Manival, Marc Quinternet. Optimizing the First TPR Domain of the Human SPAG1 Protein Provides Insight into the HSP70 and HSP90 Binding Properties. *Biochemistry*, 2021, 60 (30), pp.2349-2363. 10.1021/acs.biochem.1c00052 . hal-03175986

**HAL Id: hal-03175986**

**<https://hal.univ-lorraine.fr/hal-03175986>**

Submitted on 8 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Optimizing the first TPR domain of the human SPAG1 protein bring insight into HSP70 and HSP90 binding properties

## Authors:

Sana Dermouche,<sup>1,#</sup> Marie-Eve Chagot,<sup>1,#</sup> Xavier Manival,<sup>1</sup> Marc Quinternet <sup>2,\*</sup>

## Affiliations:

1. Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France

2. Université de Lorraine, CNRS, INSERM, IBSLor, F-54000 Nancy, France

#. Authors contributed equally.

\*. Corresponding author

## **KEYWORDS**

Protein engineering; TPR; SPAG1; R2SP; RUVBL1; RUVBL2; cilia; dynein; DNAAF; HSP90; HSP70; protein chaperone; assembly factors; ITC; NMR; DLS

## **ABSTRACT**

Tetratricopeptide Repeat domains, or TPR domains, are protein domains that mediate protein:protein interaction. As they allow contacts between proteins, they are of particular interest in transient steps of the assembly process of macromolecular complexes, such as the ribosome or the dynein arms. In this study, we focused on the first TPR domain of the human SPAG1 protein. SPAG1 is a multidomain protein important for ciliogenesis whose known mutations are linked to the primary ciliary dyskinesia syndrome. It can interact with the chaperones RUVBL1/2, HSP70 and HSP90. Using protein sequence optimization in combination with structural and biophysics approaches, we analyzed, with an atomistic precision, how the C-terminal tails of HSPs bind a variant form of SPAG1-TPR1 that mimics the wild-type domain. We discussed our results in regard of others complex 3D structures with the aim to highlight the motifs in the TPR sequences that could drive the positioning of the HSP peptides. These data could be of importance for the druggability of TPR regulators.

## INTRODUCTION

To ensure the efficient functioning of biological pathways in cells, proteins must interact through a complex network weaved within and between cellular compartments, as well as with lipid membranes and the extracellular medium. To transmit information, proteins are able of direct contacts with other proteins, nucleic acids, lipids, or small molecules. Thus, domains dedicated to the binding of specific types of partners have appeared in proteins, sometimes in several copies, to finely regulate their communication abilities <sup>1</sup>. The rise of structural biology highlighted the three-dimensional structures of many protein domains, commonly found in proteins, like domains SH2/SH3 (Src Homology 2/3), PDZ (PSD-95 Dlg1 ZO-1), Zinc fingers, RRM (RNA Recognition Motif), CRD (Carbohydrate Recognition Domain) and many more. Fascinatingly, structural biology permits to access, with an atomic precision, the molecular determinants that drive interactions between these domains and their biological partners. This latter point is a key point to investigate the druggability of protein targets, especially for proteins coming from pathogenic bacteria, viruses, and parasites <sup>2</sup>. To answer functional questions, structural biology studies require significant amounts of samples and their analyses in vitro through methods such as nuclear magnetic resonance (NMR), X-ray crystallography or cryo-electron microscopy. At a first approach, proteins of interest are purified using bacterial expression systems since they are, most of the time, easy to set up. Unfortunately, not all proteins can be efficiently purified due to their intrinsic features which can make them insoluble for example. To get through this bottleneck, prokaryotic, eukaryotic and in vitro expression systems were developed in combination with more and more complex purification protocols <sup>3-5</sup>. Another possibility is to slightly modify the wild-type sequence of a protein to facilitate expression in simple experimental conditions <sup>6</sup>. This approach requires control analyses that demonstrate that the behavior of the modified protein (meaning enzymatic activity, binding capacities, conformational adaptability, and so on) does not differ from its native form.

Extensive studies are made on the TPR domain (Tetratricopeptide Repeat) that is present in dozens of entries in the Protein Data Bank, a worldwide library that collects 3D structures of biological macromolecules. The TPR domain consists in a repetition of at least two TPR motifs, each motif being generally made of 34 amino acids shaped into a helix-turn-helix structure <sup>7</sup>. Moreover, solubilization of the domain is facilitated by a C-terminal  $\alpha$ -helix that caps hydrophobic residues of the last TPR motif <sup>7,8</sup>. To maintain the characteristic helix-turn-helix 3D shape, positions need to be conserved within the amino

acid sequence, especially for aromatic residues <sup>7</sup>. Thus, based on sequence conservation and known 3D structures, a consensus TPR motif could have been successfully designed <sup>9</sup>. The 3D features of the TPR motif give to the TPR domain the shape of a concave cradle <sup>7</sup>, until forming a superhelix in case of a long juxtaposition of many copies of TPR motifs <sup>10</sup>. From a functional point of view, the TPR domain mediates protein:protein interactions, in particular with the heat shock proteins HSP70 and HSP90. More precisely, this latter interaction is allowed thanks to five conserved polar residues located in the cradle of the TPR domain and that promote polar contacts with the four last EEVD residues of the HSPs <sup>11,12</sup>. The five conserved residues in the TPR domain form a clamp that traps the two carboxylate groups of the C-terminal aspartic acids in HSPs, hence the name of “carboxylate-clamp” given to this five residues motif <sup>13</sup>. In addition, the interaction is also maintained by hydrophobic contacts involving the penultimate valine in HSPs and hydrophobic residues in the TPR. Finally, residues upstream the EEVD sequence were also shown important in the binding. A methionine, only found in HSP90, and an isoleucine, only found in HSP70, are both located just before the EEVD motif and can accommodate a pocket at the surface of the TPR <sup>13–16</sup>. It was proposed that this Met/Ile switch represents a crucial point for the HSP selectivity <sup>13</sup>. Moreover, the amino acids upstream the Met/Ile residue can also be crucial to strengthen the binding of HSPs to a TPR domain <sup>14,15</sup>. Nonetheless, not only the cradle in the TPR is able to promote direct contacts with protein partners. Indeed, the opposite face, that displays a convex curvature, can interact with partner for regulation. It is notably the case of TPR domains in the proteins Sti1 and Sec72 <sup>17,18</sup>.

SPAG1 (Sperm-associated antigen 1) is a human protein, present in lung and testis, that possesses HSP binding abilities and that is required in the preassembly of dynein arms <sup>15,19,20</sup>. Mutations in the SPAG1 gene were shown responsible of a human syndrome, called primary ciliary dyskinesia (PCD), that affects lung and fertility functions by reducing the motility skills of flagellar and ciliary cells <sup>20</sup>. SPAG1 is a 104 kDa protein that exhibits at least four distinct domains, including three carboxylate-clamp TPR domains in a row <sup>15,21</sup>. SPAG1 is homologous to RPAP3, the two proteins possessing a RPAP3C domain that signals their interaction capacity with two other chaperones, RUVBL1 and RUVBL2 <sup>22,23</sup>. We and others extensively studied RPAP3 and SPAG1 and their interactions with HSPs and RUVBLs in the context of the assembly processes of the small nucleolar ribonucleoproteins and of dynein arms <sup>14,15,24–28</sup>. It has been proposed that HSPs carried out protein clients to their final destinations in these assembly pathways. For that purpose, HSPs interacts with the so-called R2TP and R2SP complexes that both

contained the RUVBL1 and RUVBL2 chaperones in combination, either with the RPAP3/PIH1D1 or SPAG1/PIH1D2 heterodimers, the PIH containing partner being also able to directly bind their own partners<sup>22,29–34</sup>.

Previously, we demonstrated that first and third TPR domains in SPAG1 could recruit HSPs with similar affinities, whereas the second TPR was most likely unavailable for HSP binding, certainly because of its inner hydrophobicity. We deciphered with an atomic resolution the binding mode of the third TPR domain to HSP peptides and questioned the direct GTPase activity that was initially proposed<sup>15,21</sup>. In this study, we used protein engineering to solubilize the first TPR domain of SPAG1. We showed that an optimized TPR1 could mimic the wild-type version and obtained, thanks to NMR, the precision required to finely inspect its interaction with the HSP C-terminal tails. Interestingly, TPR1 and TPR3 in SPAG1 display a distinct binding mode, especially for the HSP90 peptide. Our data are discussed in regard of the emerging interest for druggable TPR domains, in regard of known 3D structures and in regard of the functionality of the TPR domains in SPAG1.

## MATERIALS AND METHODS

### *Homology modelling*

The wild-type amino acid sequence of the human SPAG1 protein was retrieved from the UniProt database (code: Q07617). The region 206 to 327 (SPAG1-TPR1) was used for homology modelling with the I-TASSER webserver (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>)<sup>35</sup>.

### *Optimization of protein sequences*

The best 3D model of SPAG1-TPR1 provided by I-TASSER was used as entry in the PROSS webserver (<https://pross.weizmann.ac.il/step/pross-terms/>)<sup>6</sup>. Default parameters were used for protein sequence optimization.

### *Protein expression and purification*

The optimized sequences of SPAG1-TPR1 was introduced into a pNEA-3CH plasmid dedicated to the overexpression of 6xHIS tagged recombinant proteins<sup>36</sup>. Briefly, a Ca<sup>2+</sup> competent BL21(DE3) pRARE2 strain of *Escherichia coli* was transformed with the protein expression vector using a heat-shock at 42 °C. Positive clones were selected on LB-agar plates supplemented with chloramphenicol and ampicillin. An isolated colony was grown in selective LB medium at 37 °C, overnight and under agitation. 1 L of selective LB medium was seeded with the preculture and placed at 37 °C under agitation until optical density (OD) at 600 nm reached ~ 0.6. Overexpression of the recombinant protein was induced by addition of IPTG (Isopropyl β-D-1-thiogalactopyranoside) at a final concentration of 0.2 mM. The culture was placed at 20 °C, overnight and under agitation.

Then, bacteria were harvested by centrifugation. The pellet was resuspended in lysis buffer (25 mM HEPES, pH 7.5, 300 mM NaCl, 0.5 mM TCEP, 10 mM Imidazole) and sonicated. After centrifugation, the supernatant was put on TALON beads for 2 hours, at 4 °C and under agitation. The beads were washed 3 times with the lysis buffer. The elution of bound proteins was performed in lysis buffer by adding recombinant 3C PreScission protease directly on the beads. After cleavage, the non-native GPHM sequence remains at the N-terminal extremity of the recombinant protein. Finally, a gel filtration (GF) was performed in 10 mM NaPi, pH 6.4, 150 mM NaCl, 0.5 mM TCEP with a Superdex-75 column. The GF fractions containing the protein of interest were pooled and concentrated up to a protein

concentration of 1 mM. The purification process was followed with denaturant poly-acrylamide gel electrophoresis (SDS-PAGE).

The same protocol was applied to obtain samples suitable for NMR spectroscopy, except that bacteria were grown in a minimal M9 medium supplemented with  $^{15}\text{NH}_4\text{Cl}$  and  $^{12}\text{C}$  or  $^{13}\text{C}$ -d6-glucose.

#### *Dynamic light scattering (DLS)*

DLS experiments were performed at 20 °C, in 10 mM NaPi, pH 6.4, 150 mM NaCl, 0.5 mM TCEP, on a Zetasizer Nano ZS (Malvern Panalytical) in a low-volume quartz cell. Data were processed with the Zetasizer software (Malvern) with default parameters.

#### *Protein co-expression assays*

Sequences of the human HSP70 (uniprot ID: P0DMV8) and HSP90-MC (N domain truncated HSP90, uniprot ID: P08238) proteins and of SPAG1-TPR1<sub>opt</sub> were respectively introduced in pnYK (for untagged proteins) and pnEA-3CH (for 6xHIS tagged proteins) vectors.  $\text{Ca}^{2+}$  competent BL21(DE3) pRARE2 strain of *Escherichia coli* were simultaneously transformed with the pnYK and pnEA vectors. Control experiments were done with pnYK-transformed bacteria. Our protocol was previously described <sup>15</sup>. Briefly, positive clones were selected on LB-agar plates, picked for pre-culture at 37 °C in selective LB medium and grown in LB medium at 37 °C until OD at 600 nm reached ~ 0.6. Overexpression of recombinant proteins was induced by 0.2 mM IPTG for one night at 20 °C. Cells were harvested, pelleted, and sonicated in the lysis buffer described above. Sonicates were centrifuged and the supernatant was placed on TALON beads for 2 hours at 4 °C under agitation. Proteins were eluted from the beads using lysis buffer supplemented with 300 mM Imidazole. The purification process was followed on denaturant poly-acrylamide gels.

#### *Isothermal titration calorimetry (ITC)*

The binding of synthetic peptides DASRMEEVD (C-terminal amino acids of human HSP90) and SGPTIEEVD (C-terminal amino acids of human HSP70) to recombinant SPAG1-TPR1<sub>opt</sub> was measured at 283 K on an ITC-200 microcalorimeter (Malvern) in 10 mM NaPi, pH 6.4, 150 mM NaCl, TCEP 0.5 mM. Peptide and protein concentrations were comprised between 1.5 and 2 mM and 0.1 and 0.2 mM, respectively. Peptides and proteins were dispatched into the syringe and cell, respectively. The

dissociation constant (Kd) the stoichiometry (N) and the variations in enthalpy ( $\Delta H$ ), entropy ( $\Delta S$ ) and free energy ( $\Delta G$ ) were derived from the calorimetric data using the ORIGIN7 software (Malvern).

### *Nuclear Magnetic Resonance*

NMR data were recorded at 293 K on a  $^{13}\text{C}/^{15}\text{N}$  labelled sample of SPAG1-TPR1<sub>opt</sub> concentrated at 1 mM in 10 mM NaPi, pH 6.4, 150 mM NaCl, 0.5 mM TCEP using a 600 MHz spectrometer equipped with a cryoprobe. Usual 2D and 3D NMR spectra were collected. The majority of  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  resonances were assigned and deposited in the Biological Magnetic Resonance Bank under access code 34586.

The 3D NMR structure of SPAG1-TPR1<sub>opt</sub> was calculated using distance restraints derived from  $^1\text{H}$ - $^{15}\text{N}$  3D-NOESYHSQC,  $^1\text{H}$ - $^{13}\text{C}$  3D-NOESYHSQC and  $^1\text{H}$ - $^1\text{H}$  2D-NOESY experiments recorded with a mixing time of 120 ms, and dihedral restraints derived from TALOS-N and PREDITOR<sup>37,38</sup>. Automated procedure of CYANA 3.0 was used to iteratively assign NOE cross-peaks<sup>39</sup>. After careful inspection of the final CYANA step, the distance and dihedral restraints were used to calculate 100 water-refined structures with the AMBER web portal<sup>40</sup>. The 20 structures with the lowest restraint energy were selected as representative structures of SPAG1-TPR1<sub>opt</sub> and deposited in the Protein Data Bank under access code 7BEV.

Samples containing 1 mM of  $^{13}\text{C}/^{15}\text{N}$  labelled SPAG1-TPR1<sub>opt</sub> and 3 mM of DASRMEEVD or SGPTIEEVD peptides were prepared in 10 mM NaPi, pH 6.4, 150 mM NaCl, 0.5 mM TCEP. In addition to classical 2D and 3D NMR spectra, 2D and 3D filtered NOESY with a mixing time of 120 ms data were collected at 600 MHz and 293K, to highlight intermolecular cross peaks. To evaluate the chemical shift perturbations (CSP) induced by HSP peptide binding to labelled SPAG1-TPR1<sub>opt</sub>, the variation of composite  $^1\text{H}$ - $^{15}\text{N}$  chemical shift was calculated, for each backbone amide group, as follow:  $\Delta\delta = \sqrt{0.1(\delta^{15}\text{N})^2 + (\delta^1\text{H})^2}$  (where  $\delta^{15}\text{N}$  and  $\delta^1\text{H}$  are respectively the chemical shift difference in the  $^{15}\text{N}$  dimension and in the  $^1\text{H}$  dimension between the free and the bound states of SPAG1-TPR1<sub>opt</sub>).

### *Modelling of the peptide:SPAG1-TPR1<sub>opt</sub> complexes*

We used the best NMR structure of SPAG1-TPR1<sub>opt</sub> and ten structures of the DASRMEEVD and SGPTIEEVD peptides to model complex structures using HADDOCK-2.2 and CNS-1.3<sup>41</sup>. 3D structures of the two peptides were extracted from the 3D NMR structure of RPAP3-TPR2 in complex with DTSRMEEVD and SGPTIEEVD peptides. The threonine in the HSP90 peptide was mutated into alanine



using PyMOL <sup>42</sup>. To drive the docking, we defined active and passive residues to build ambiguous intermolecular restraints (AIR). In SPAG1-TPR1<sub>opt</sub>, active residues are solvent exposed residues that (i), are located on the concave face of structure and (ii), that belong to the 25 % most perturbed residues in NMR titration experiments (i.e with  $\Delta\delta$  values > centile 75 value). Passive residues are surface residues neighbors of the active residues. All residues in the peptides were defined as active. Intermolecular NOEs were turned into unambiguous restraints. Additional unambiguous distance restraints were added between residues of the carboxylate-clamp in SPAG1-TPR1<sub>opt</sub> and the C-terminal part of the peptides. Automated mode was chosen for the definition of the semi-flexible interfaces. 2560 and 256 structures were calculated in the it0 rigid-body and the it1 semi-flexible simulated annealing steps, respectively. The last 256 structures were refined in water. The docking solutions were clustered using the FCC (Fraction of Common Contacts) algorithm with a cutoff value of 0.75 <sup>43</sup>. The 3D structures of the complexes were ranked according to their HADDOCK score.

## RESULTS

### *Sequence optimization of SPAG1-TPR1 greatly enhances its solubility*

Previously, we showed that the human SPAG1 protein contained three TPR domains<sup>15</sup>. When produced in an isolated form, two of them were soluble, meaning the first and the third domains, the second domain being not soluble. Unfortunately, due to aggregation features appearing at a protein concentration greater than 150  $\mu$ M, we were not able to perform an in-depth analysis of SPAG1-TPR1, which has however been shown able to bind HSP70 and HSP90 through their C-terminal tail<sup>15</sup>. Indeed, the wild-type SPAG1-TPR1 (SPAG1-TPR1<sub>wt</sub>) possesses residues K213, N217, N247, K277 and R281 that form a conserved carboxylate-clamp motif, which is the signature for the binding of HSP70 and HSP90 tails. Thus, we thought that study at the atomic level could bring new insights into the molecular properties of this protein domain.

As the TPR domain is a well-characterized domain in the literature, we used the I-TASSER homology modelling webserver to have a first, but potentially strong, preview of the 3D structure of SPAG-TPR1<sub>wt</sub><sup>35</sup>. Using the amino-acid sequence of region 206-327 of the human SPAG1 protein as the sole source of information, we got a 3D model with a C-score value of 0.66 (Figures 1A, 1B and 1C). This value is sufficiently high to consider the modelling as reliable. Interestingly, the best threading 3D template was SPAG1-TPR3 (PDB code: 6I57). As expected, the prediction consists in seven  $\alpha$ -helices folded into three TPR motifs followed by a final  $\alpha$ -helix that caps the sixth  $\alpha$ -helix (Figure 1A). The global shape of the protein domain displays a convex face and a concave face, where carboxylate-clamp residues are available for HSP binding. Analysis of the predicted structure revealed solvent-exposed hydrophobic amino acids, especially residues F<sub>207</sub>, L<sub>208</sub>, L<sub>239</sub>, V<sub>243</sub>, L<sub>255</sub>, F<sub>262</sub>, L<sub>271</sub>, V<sub>305</sub>, L<sub>311</sub> and L<sub>323</sub>, that could promote protein aggregation through uncontrolled intermolecular hydrophobic contacts (Figure 1B).

Next, considering this prediction, we used the PROSS webserver with the idea to design an optimized sequence suitable for recombinant expression of SPAG1-TPR1 in *E. coli*<sup>6</sup>. For this purpose, PROSS requires a 3D structure and an amino acid sequence alignment. Based on the I-TASSER 3D structure of SPAG1-TPR1<sub>wt</sub>, the PROSS algorithm returned seven primary sequences, starting with 12 mutations up to 40 mutations with respect to the wild-type sequence (Figure 1C). We previously demonstrated that not only  $\alpha$ -helices 1, 3 and 5 (that hold the carboxylate-clamp residues) but also the region at the root of  $\alpha$ -helix 7 could be involved in the recognition of HSP peptides in RPAP3-TPR2 or SPAG1-TPR3<sup>14,15</sup>. In the light of these data, we rationally designed an amino acid sequence of SPAG1-TPR1 based on the

PROSS proposals, but only for residues distant from the carboxylate-clamp residues and located on the convex face of the domain. In addition, we tried to restrict the number of mutations introduced in the wild-type sequence to minimize their potential impact on the global features of the protein domain. Thus, we constructed a final sequence with the following seven mutations: F207Y, E252Y, F262E, T285A, T296R, S300K and D304K (Figure 1C). This optimized sequence was introduced in a plasmid dedicated to protein overexpression in bacteria. We showed that our optimized SPAG1-TPR1 (SPAG1-TPR1<sub>opt</sub>) was efficiently purified from *E. coli* (Figure 1D). Interestingly, with a mean size of  $4.6 \pm 0.1$  nm, dynamic light scattering (DLS) showed that this recombinant domain remains monomeric in solution at protein concentration of 1 mM (Figure 1E). This result strongly contrasts with the DLS experiments performed on SPAG1-TPR1<sub>wt</sub> (i.e. wild-type SPAG1-TPR1), for which a mean size of  $\sim 80$  nm was previously measured at protein concentration greater than  $150 \mu\text{M}$ ,<sup>15</sup> and demonstrates the stabilization effect provided by the point mutations.

#### *Optimized and wild-type SPAG1-TPR1 behave similarly*

Thanks to protein sequence optimization, we obtained a highly soluble sample of SPAG1-TPR1. However, one could wonder whether this domain could reproduce the behavior of the wild-type domain. To answer this question, we first performed protein co-expression assays between His-tagged SPAG1-TPR1<sub>opt</sub> and human HSP70 or HSP90-MC (i.e. HSP90 truncated of its N-terminal domain). As observed for the wild-type SPAG1-TPR1<sub>wt</sub>, we showed that SPAG1-TPR1<sub>opt</sub> co-elutes specifically with HSPs (Figure 2A). We next performed isothermal titration calorimetry (ITC) experiments to access the thermodynamics of the interaction between SPAG1-TPR1<sub>opt</sub> and HSP peptides (DASRMEEVD for HSP90 and SGPTIEEVD for HSP70) (Figure 2B). For the two peptides, we obtained stoichiometry values close to 1:1. The dissociation constant (K<sub>d</sub>) values were  $2.1 \pm 0.2 \mu\text{M}$  and  $11.0 \pm 0.9 \mu\text{M}$  for HSP90 and HSP70 respectively. These ITC data were similar to the ITC data obtained with SPAG1-TPR1<sub>wt</sub>, for which we measured, in the same experimental conditions, 1:1 stoichiometries and K<sub>d</sub> values of  $2.5 \pm 0.2 \mu\text{M}$  and  $15.8 \pm 2.6 \mu\text{M}$  for HSP90 and HSP70 peptides<sup>15</sup>. More, for each peptide, values of variations in enthalpy ( $\Delta H$ ), entropy ( $\Delta S$ ) and free energy ( $\Delta G$ ) are almost identical whatever the domain tested<sup>15</sup>. Altogether, these biochemical and biophysical data strongly support the fact that optimized and wild-type SPAG1-TPR1s behave similarly in vitro. Thus, they allowed us to go deeper in the structural analysis of SPAG1-TPR1.

### *Structural analysis of SPAG1-TPR1<sub>opt</sub>*

We used nuclear magnetic resonance in solution to get experimental data at the atomic scale on SPAG1-TPR1<sub>opt</sub>. We produced a <sup>13</sup>C/<sup>15</sup>N labelled of SPAG1-TPR1<sub>opt</sub> concentrated at 1 mM and recorded a set of 2D and 3D spectra to get its resonance assignment.

In agreement with DLS data, the <sup>1</sup>H-<sup>15</sup>N HSQC spectrum displays the features of a monomeric and well folded protein domain. Indeed, the cross peaks observed on this 2D spectrum are thin, with homogeneous intensities and well dispersed in the two dimensions (Figure 3A). Analysis of the 3D spectra permitted to assign the large majority of <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N resonances of SPAG1-TPR1<sub>opt</sub> and allowed us to access the 3D structure of the protein domain. We obtained a well-defined ensemble of twenty 3D structures of SPAG1-TPR1<sub>opt</sub> (for statistics, see Table 1). Our experimental structure reveals that SPAG1-TPR1<sub>opt</sub> folds into a TPR domain with a seventh capping  $\alpha$ -helix (Figure 3B), as supported by the HSP binding assays. Here again, the concave face, which is made of  $\alpha$ -helices 1, 3, 5 and 7, exposes the five residues of the carboxylate-clamp, whereas a convex face is made of  $\alpha$ -helices 2, 4 and 6. With a RMSD value of 0.9 Å measured on C $\alpha$  atoms, this structure superimposed well with the I-TASSER model generated with the wild-type sequence of SPAG1-TPR1 (Figure 3C).

Next, we naturally focused on the optimized positions to better understand their role in the stabilization of the domain. For position 207, the tyrosine residue, which is quite solvent exposed, bring polarity to the side chain without changing the hydrophobic nature of the phenol ring coming from the initial phenylalanine. Positions 252 and 262 were of particular interest since our mutations turned drastically the chemical feature of the amino acid side chain. In SPAG1-TPR1<sub>opt</sub>, Y252 is framed by V229, L255 and R248 (Figure 3D). Moreover, its hydroxyl group brings surface polarity. At the opposite, E262 in SPAG1-TPR1<sub>opt</sub> promotes stable polar contacts with N259, Q263, Q289 and K291. Obviously, this polar environment appeared not favorable to host the phenylalanine present in the wild-type sequence. In position 285, the conversion of the threonine to an alanine residue, in a hydrophobic context supported by I253, W258 and Y286, cancels the polar feature of the hydroxyl group. Concerning position 296, introducing an arginine residue permitted (*i*), to efficiently connect helices  $\alpha$ 6 and  $\alpha$ 7 through a salt bridge with E320, and (*ii*), to limit exposition to solvent of L316 and L323. Finally, lysine residues introduced in position 300 and 304 enhances the general solubility thanks to their side chain amine group. Thus, our NMR structure fully explains the high stability of SPAG1-TPR1<sub>opt</sub>.

*SPAG1-TPR1<sub>opt</sub> does not bind HSP C-terminal tails as SPAG1-TPR3 does*

Having solved the 3D structure in solution of SPAG1-TPR1<sub>opt</sub> and demonstrated that it could nicely mimic the wild-type domain, we aimed to have more precise information on the interaction with HSP90 and HSP70 C-terminal tails.

To this end, we mapped the footprint of HSP C-terminal tails on the SPAG1-TPR1<sub>opt</sub> 3D structure by analyzing the perturbations of the <sup>1</sup>H-<sup>15</sup>N composite chemical shifts induced upon formation of the protein:peptide complex (CSP, for chemical shift perturbation). This analysis revealed that,  $\alpha$ -helices 1, 3 and 5 as well as the  $\alpha$ 6- $\alpha$ 7 bridging region at the root of the last  $\alpha$ -helix in SPAG1-TPR1<sub>opt</sub> appear involved in the binding of the two peptides (Figure 4). Analysis of filtered NOESY spectra revealed a common pattern of intermolecular NOEs in the complexes with the two peptides (Figure 4). Indeed, Met in HSP90, but also Ile in HSP70, faced I253, T284 and L311 in SPAG1. Ala in HSP90 was found close to N309 region in SPAG1, while no intermolecular NOE could have been highlighted in this region with the HSP70 peptide (Figure 4). These NMR data reminded us the binding features observed in HSP-bound RPAP3-TPR2<sup>14</sup>. Indeed, in addition to comparable CSP profiles obtained upon HSP peptides binding, RPAP3-TPR2 and SPAG1-TPR1<sub>opt</sub> share patterns of intermolecular NOEs involving conserved positions on the peptides and on the protein domains (supplemental figure S1). This strongly suggests that HSP peptides adopt similar conformations upon binding of RPAP3-TPR2 and SPAG1-TPR1<sub>opt</sub>.

To have 3D views of these interactions, we modeled the protein:peptide complexes in a HADDOCK protocol. Thus, as entries, we used the best SPAG1-TPR1<sub>opt</sub> NMR structure and ten distinct structures of the HSP peptides extracted from the complexes with RPAP3-TPR2 (PDB codes: 6FDT and 6FDP). To drive the docking, we used our ambiguous and unambiguous restraints derived from NMR data. To ensure the orientation of the peptide, we added distances locking the EEVD parts of the peptides next to the conserved carboxylate-clamp region that were shown essential in TPR domains for efficient HSP binding (supplemental figure S2)<sup>13,14,16,27</sup>. The results of the calculations were summarized in Table 2.

We obtained, for each complex, final 3D solutions that all cluster in one group, meaning that all the structures share 75 % of common contacts. For the two complexes, the five best HADDOCK structures superimposed well, and the positioning of the peptides, that both adopt a horseshoe conformation, nicely matches with the NMR CSP footprints and intermolecular NOEs (Figures 5A and 5B). Due to their importance in the cochaperone selectivity, we focused our intention on the Met and Ile residues in the

HSP peptides. These 3D models reveal that these two residues are both hosted by residues Q250, I253, L280, T284 and L311. We compared our 3D data with those we obtained with the third TPR in SPAG1 known to be the other module able to bind HSPs in the full-length protein.

We noticed a clear difference in the positioning of the Met side chain in the HSP90 peptide (Figure 5C). The T284L substitution in SPAG1-TPR3 may prevent, by steric hindrance, a deep hosting of the Met methyl group close to this position. The Y residue in SPAG1-TRP3, replacing L280 in SPAG1-TRP1<sub>opt</sub>, seems to act as a barrier, which, in addition to the L311E and N309I substitutions that displace hydrophobicity at the bottom of SPAG1-TPR3, promotes opposite orientations of the Met side chain in the two complexes. Concerning the HSP70 peptide, its Ile residue is bulkier, more hydrophobic, and less plastic than the Met in HSP90. The L280 in SPAG1-TPR1<sub>opt</sub> leaves enough space to host the two methyl groups of Ile in a quite hydrophobic environment spreading from T284 to V276 (Figure 5C). For that reason, the positioning of Ile in the two complexes is not as different as in the case of the Met in HSP90. Thus, even if TPR 1 and 3 recruit HSPs with similar affinities, they take distinct structural paths to bind their C-terminal ends, especially for HSP90.

## DISCUSSION

### *Optimization of the TPR domains*

In vitro techniques used for atomistic structural studies are usually dependent on the quality of the samples. De facto, not all proteins can be analyzed due to a misbehavior, according to the experimenter's expectations, during the expression or purification process. Here, we used protein engineering to solubilize TPR 1 of SPAG1.

The TPR domain is a versatile domain that is highly studied since it mediates protein/protein interaction, especially with HSPs if the carboxylate-clamp signature is present in the sequence. The amino acids surrounding the structuring residues required for establishing the TPR motif, could display high heterogeneity<sup>7,9</sup>. As we demonstrated, this heterogeneity of sequence could lead to protein insolubility in SPAG1-TPR2<sup>15</sup>. By mutating less than 6 % of the amino acid sequences, we obtained a soluble and highly stabilized SPAG1-TPR1 in vitro. Remarkably, the SPAG1-TPR1<sub>opt</sub> sample was stable for more than one year at 4 °C in the NMR buffer. This demonstrates that clever modifications in TPR domains permit their structural study alone or in complex with biological partners.

One great interest of the optimization approach is to support studies on the druggability of proteins. Indeed, in silico methods are often used to screen plethora of ligands toward protein targets<sup>44</sup>. Robust 3D structure of the target is required for an efficient screening. Enhancing the solubility of the target will facilitate resolution of structures, which can directly be used in bioinformatic binding assays, or which can first serve as starting template for homology modelling of the wild-type protein. It also permits bench scale experiments, this step being crucial to test the real binding activity of computational ligands. Moreover, stabilized samples appear also essential to test the impact of the ligand on the activity of the target, in vitro. Interestingly, using combined in silico and biochemical approaches, the TPR2A domain of the protein Hop was turned into a druggable target<sup>45</sup>. Indeed, authors designed an organic ligand that binds the HSP90 recognition site through contacts with the carboxylate-clamp residues. Similarly, Dictyoceratin-A and -C are two inhibitor molecules that bind the TPR1 region of RPAP3<sup>46</sup>. These studies open the door to a fine tuning of the HSP90 activity via the selective blocking of its different adaptative regulators. Our novel data on SPAG1-TPR1<sub>opt</sub> in complex with DASRMEEVD and SGPTIEEVD peptides extend the knowledge on the binding mode of TPR domains containing a carboxylate-clamp signature and could contribute to the development of novel molecular compounds with specific modulator activity on HSP70 and/or HSP90 co-chaperones.

Since fine tuning of the HSP activity goes hand in hand with fine knowledge of the interaction modes of the co-chaperones, we compared our 3D data with known structures available in the Protein Data Bank and in our previous studies.

*Sequence signature in the TPR domain might forecast the mode of binding of the last Methionine in HSP90*

Concerning the interaction between the C-terminal tail of HSP90 and TPR domains, we observed that the Methionine of the MEEVD motif could adopt 4 main orientations, defining four families of structures (Figures 6A and 6B). In one of them, the Met residue points toward the solvent and does not seem involved in the anchoring of the peptide onto the TPR domain (Family 0, Figure 6A). The three remaining groups define three distinct interaction spots for the methionine at the surface of the TPR domains (Figures 6B and 6C). On a structural alignment we looked at primary sequences in these patches with the aim to highlight residues that could be the signature of one or the other interaction modes.

The family 1 of structures (Figure 6A) contains the first ever described 3D structure of an HSP90 peptide in complex with a TPR domain<sup>13</sup>. In this group, the peptide is quite linear, spreading from the top to the bottom of the TPR surface. It distinguishes this family from families 2 and 3 in which the peptide adopts superimposable conformations looking like a horseshoe, extended from the right to the left of the TPR (Figure 6A). For TPR domain orientation, please report to Figure 6. Remarkably, the HSP90 peptides used in the structures of the family 1 is very short, as it contains only the five last MEEVD residues. In contrary, for 3D determination of structures included in the families 2 and 3, longer peptide sequences were used. The interaction spot in family 1 (spot 1) seems quite well conserved in families 2 and 3 (Figure 6C), making difficult to highlight specific positions for family 1. Thus, one could wonder how the length of the peptide influences its positioning. The interaction spot 2 in family 2 looks quite heterogenous in terms of amino acid sequences, corresponding residues in families 1 and 3 reproduce this heterogeneity. Nonetheless, our detailed inspection of these structural data drawn our attention to position 3 in the interaction spot 3, where a Thr residue is strictly conserved in structures belonging to family 3. This Thr is replaced by a Asn in family 1 and, in family 2, by bulkier residues (Leu or Lys) or by a Ser, which is more hydrophilic than Thr. This spot 3, quite hydrophobic in family 3, appears more susceptible to host the methyl group of the HSP90 methionine than in the other families. From our analysis, we propose that this Thr residue, which corresponds to Thr284 in optimized SPAG1-TPR1<sub>opt</sub>,



may be a key residue that signals the positioning of the HSP90 methionine close to it. TPR domains that lacks this Thr could be direct to other families, but clear sequence determinants still escape our analysis. To go further, we performed the same structural analysis for TPR domains in complex with the HSP70 C-terminal tail. In the 3D structures retrieved from the Protein Data bank, the HSP70 peptides include an Ile or a Val residue just before the EEVD end. These residues are bulkier and largely more hydrophobic than the methionine of the HSP90 end. As we showed with the comparison of 3D data obtained on TPR1<sub>opt</sub> and TPR3 (Figure 5), accommodation of the Ile/Val side chains to the molecular surface of the TPR domain requires space and enough hydrophobicity to minimize exposition with the solvent. This point is even more important since the amino acids located before the Ile/Val residue, meaning Ser, Gly, Pro and Thr, do not possess the same interaction power than the corresponding Asp, Ala, Ser and Arg in HSP90. Thus, compared to the DASR sequence in HSP90, the SGPT sequence in HSP70 is likely less susceptible to bind an anchoring point at the surface of the TPR domain. We classified the 3D structures in two families (Figure 7). Family 1 is close to the family 1 defined with the HSP90 peptides, since the HSP70 peptide is linear and spreads from the top to the bottom of the protein receptor. The Val/Ile side chains are in regard of the top right corner of the TPR. In family 2, Val or Ile are in the bottom left corner. For TPR domain orientation, please report to Figure 7. We compared TPR sequences in these two spots. Within the family 1, spot 1 displays heterogeneity in terms of amino acid content. At this spot, family 2 of structures exhibit also inhomogeneous sequences. The same comments could be done for spot 2, where various amino acids occupy each position in the two families of structures. Based on our structural alignment, and to our mind, it was not possible to highlight sequence features that could predict the positioning of the Ile residue in the HSP70 peptide.

All these observations can be related to the lower affinity that is generally observed for HSP70 than HSP90 peptide for a given TPR domain. Accommodation of residues upstream the EEVD end to the TPR domain is more problematic for the HSP70 peptide that does not displays highly attractive chemical functions within its first residues. We think that the N-terminal end of the HSP70 peptide adapts the TPR surface to minimize exposition of its Ile residue to the solvent. However, having so much variability between the sequences of the TPR family is probably not the best mean to offer optimized conditions to reduce solvent exposition. In contrary, the HSP90 peptide appears more susceptible to form chemical bound thanks to its first Asp or Arg residues. In addition, due to its size smaller than Ile, its methionine

offers more possibilities to accommodate the TPR surface. Thus, the HSP90 peptide sequence seems more adapted to sample and find the most favorable energetic positions on the TPR domain.

### *HSP binding properties of SPAG1*

Previously, we described SPAG1 as a platform for HSP proteins through interactions with its TPR1 and TPR3. We demonstrated that the wild-type and optimized versions of SPAG1-TPR1 behaved very similarly toward HSP binding, especially in terms of thermodynamics, strongly suggesting that they recruit the HSP C-terminal tails with the same structural recognition mode. Despite the T285A substitution (which occurs sequentially close to, but spatially quite distant from the Thr284 that was proposed as a marker of family 3 for HSP90 peptides), we can reasonably rank SPAG1-TPR1<sub>wt</sub> into the same families that were described for SPAG1-TPR1<sub>opt</sub> (Figures 6 and 7). Thus, the structural differences raised between the HSP70- and HSP90-bound forms of SPAG1-TPR1<sub>opt</sub> and SPAG1-TPR3 could rationally be transposed to SPAG1-TPR1<sub>wt</sub>.

This result suggests that SPAG1 needs a strong partnership with HSPs. The two chaperones HSP70 and HSP90 are both able to hold and deliver client proteins in an assembly pathway. One could assume that delivering of HSP clients via SPAG1 (which works in concert with PIH1D2, RUVBL1 and RUVBL1 into a so-called R2SP complex) is a so crucial function for the cell expressing this protein that two TPR domains, with similar properties, were included in the sequence. Having two structural modes of recognition for the HSPs prevent the total blocking of the SPAG1 function if an inhibitor molecule, specific from one or the other mode, enters the cell. Another possibility could be that two HSPs could be bound to SPAG1 with the aim to bring faster, or in larger quantities, client proteins that need to be assembled in a specific pathway. It could also facilitate the exchange of clients between HSP70 and HSP90 if the two chaperones bind SPAG1 simultaneously.

## **ACCESSION CODES**

The NMR structure and the  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts of SPAG1-TPR1<sub>opt</sub> were deposited in the Protein Data Bank and in the Biological Magnetic Resonance Bank under access codes 7BEV and 34586 respectively. Uniprot accession codes of the human proteins SPAG1, HSP70 and HSP90 used in this study are respectively, Q07617, P0DMV8 and P08238.

## **ACKNOWLEDGMENTS**

This work was supported by the French National Center for Scientific Research (CNRS), Université de Lorraine (France), the 'Région Grand Est' and the 'Agence Nationale de la Recherche' [Grant No. ANR-16-CE11-0032-02]. Access to the NMR, DLS and ITC core facilities of the Biophysics and Structural Biology department of UMS2008/US40 IBSLor was deeply appreciated (<https://umsibslor.univ-lorraine.fr/>).

## FIGURE LEGENDS

*Figure 1: Optimizing the TPR1 domain in SPAG1.*

(A) Two views of the 3D structure of region 206-327 in SPAG1 obtained by homology with the I-TASSER software. TPR motifs and the C-terminal capping helix are highlighted using different colors. (B) Hydrophobic residues in SPAG1-TPR1<sub>wt</sub> are represented using sticks on the 3D homology model. (C) Results of the PROSS optimization. Sequences of the wild-type, the PROSS proposals and the in-house optimized SPAG1-TPR1 are aligned. Strictly conserved residues are in black boxes. Residues of the carboxylate-clamp motif are highlighted with an asterisk. Secondary structures coming from the 3D homology model are displayed on the top of the alignment. On the right, carboxylate-clamp residues (in black, with an asterisk) and residues selected for optimization in SPAG1-TPR1<sub>wt</sub> are displayed using sticks on the 3D homology model. (D) Purification process of recombinant SPAG1-TPR1<sub>opt</sub>. On the SDS-PAGE profile displayed on the left, So, CSo, SSo, Unbound, Beads, MW, RAC, Wash and Cut3C refer respectively to the sonicate, the sonicate pellet, the sonicate supernatant, the flow-through of TALON beads, the TALON beads, the size marker, the TALON beads after 3C cleavage, the washed TALON beads after 3C cleavage and the soluble fraction after 3C cleavage. The protein of interest is indicated using a black arrow. The gel filtration profile was obtained on a preparative Superdex-75 column. Fractions of interest are highlighted in a grey box. The maximal intensity measured at 280 nm is indicated. The SDS-PAGE on the right corresponds to the analysis of fractions 15 to 22 coming from the gel filtration. (E) DLS analysis of SPAG1-TPR1<sub>opt</sub> at 150, 500 and 1000  $\mu$ M. The size corresponding to the highest intensity is reported.

*Figure 2: SPAG1-TPR1<sub>opt</sub> interacts with HSPs.*

(A) Co-expression assays between SPAG1-TPR1<sub>opt</sub> and HSP90-MC or HSP70. On the SDS-PAGE profile, SSo, Beads, Elu and MW refer to the supernatant sonicate, the TALON beads, the soluble fraction after elution with Imidazole buffer and to the size marker. Test experiments are on the left part of the gel, control experiments are on the right part. (B) ITC experiments recorded at 283 K with SPAG1-TPR1<sub>opt</sub> and DASRMEEVD or SGPTIEEVD peptides in 10 mM NaPi, pH 6.4, 150 mM NaCl, 0.5 mM TCEP. The protein concentration was comprised between 0.15 and 0.2 mM. Peptides were

concentrated between 1.5 and 2 mM. N, K<sub>d</sub>, T, ΔH, ΔS and ΔG refer respectively to the stoichiometry, the dissociation constant, the temperature and the variations in enthalpy, entropy, and free energy.

*Figure 3: NMR solution structure of SPAG1-TPR1<sub>opt</sub>.*

(A) <sup>1</sup>H-<sup>15</sup>N HSQC spectrum of SPAG1-TPR1<sub>opt</sub> recorded at 600 MHz, 293 K, in 10 mM NaPi, pH 6.4, 150 mM NaCl, 0.5 mM TCEP. Folded resonances were colored in blue. (B) Two backbone representations of the twenty best NMR structures after AMBER water-refinement. The three TPR motifs and the C-terminal capping helix are highlighted. (C) Superimposition of the SPAG1-TPR1<sub>opt</sub> NMR structure with the 3D homology model of SPAG1-TPR1<sub>wt</sub>. (D) Structural close-up on the regions of residues Y252, E262, R296 and A285 in SPAG1-TPR1<sub>opt</sub>.

*Figure 4: NMR analysis of SPAG1-TPR1<sub>opt</sub> upon HSP peptide binding.*

(A) Chemical shift perturbation profiles (CSP) of SPAG1-TPR1<sub>opt</sub> upon binding of the DASRMEEVD peptide. The <sup>1</sup>H-<sup>15</sup>N chemical shift variation (Δδ) for each residue is plotted against the sequence number. Positions of α helices in SPAG1-TPR1<sub>opt</sub> are represented on the diagram. The centile-75 value of Δδ, which discriminates the most perturbed residues, is indicated as a line. The residues displaying Δδ value higher than the centile-75 value are colored on a molecular surface representation of SPAG1-TPR1<sub>opt</sub>. On the right, intermolecular NOE cross peaks are represented on strips extracted from a filtered 3D <sup>1</sup>H-<sup>13</sup>C NOESYHSQC. (B) Same as (A) but for the SGPTIEEVD peptide. CSP data were recorded at 600 MHz, 293 K, in 10 mM NaPi, pH 6.4, 150 mM NaCl, 0.5 mM TCEP using a peptide:protein ratio of 3.

*Figure 5: HADDOCK modelling of SPAG1-TPR1<sub>opt</sub> in complex with HSP peptides.*

(A and B) The five best 3D models (i.e. presenting the best HADDOCK scores) of SPAG1-TPR1<sub>opt</sub> in complex with DASRMEEVD (A) and SGPTIEEVD (B) are superimposed and presented on the left. The peptide is represented using its backbone trace. On the right, detailed views of the methionine (A) and of the isoleucine (B) environments are displayed using the best HADDOCK structure of each complex. (C) Superimposition of the HSP-bound states of SPAG1-TPR1<sub>opt</sub> and of SPAG1-TPR3. For SPAG1-TPR1<sub>opt</sub>, the best HADDOCK structures were used (in blue or magenta). For SPAG1-TPR3, representative structures coming from the work of Chagot and colleagues were used (in light or dark

grey) <sup>15</sup>. Detailed views of the environment of the methionine and of the isoleucine in HSP peptides are presented. The numbering of SPAG1-TPR1<sub>opt</sub> was used. Residues in the SPAG1-TPR1<sub>opt</sub> and SPAG1-TPR3 complexes are labeled in blue/magenta and grey, respectively.

*Figure 6: Structural comparison of TPR domains in complex with HSP90 peptides*

(A) The families 0 to 3 of superimposed 3D structures are represented using the backbone traces of the TPR domain and cartoon representations of the HSP90 peptides. Arbitral orientations of the structures are specified. When available, the name of the TPR protein and the PDB code are indicated. (B) All the 3D structures displayed in (A) are superimposed and, for each of them, the methionine in the HSP90 peptide is represented using sticks. The three main spots of interaction involving this methionine are highlighted with colored circles. Two views are proposed. (C) Detailed views of the binding spots of each family. To look at sequence conservation on the structural alignment, five positions (#1 to #5) are defined at each spot. The amino acid that occupies each of these positions in all the structures is reported, structures being sorted by family. The residues belonging to the carboxylate-clamp motif are indicated with an asterisk.

*Figure 7: Structural comparison of TPR domains in complex with HSP70 peptides*

(A) The families 1 and 2 of superimposed 3D structures are represented using the backbone traces of the TPR domain and cartoon representations of the HSP70 peptides. Arbitral orientations of the structures are specified. When available, the name of the TPR protein and the PDB code are indicated. (B) All the 3D structures displayed in (A) are superimposed and, for each of them, the isoleucine, or the corresponding valine, in the HSP70 peptides is represented using sticks. The two main spots of interaction involving this isoleucine/valine are highlighted with colored circles. Two views are proposed. (C) Detailed views of the binding spots of each family. To look at sequence conservation on the structural alignment, up to eight positions (#1 to #8) are defined at each spot. The amino acid that occupies each of these positions in all the structures is reported, structures being sorted by family. The residues belonging to the carboxylate-clamp motif are indicated with an asterisk.

TABLES

<b>SPAG1-TPR1<sub>opt</sub></b>	
<b>NMR distances and dihedral constraints</b>	
Total NOEs	2919
Short range ( $ i - j  \leq 1$ )	1357
Medium-range ( $ i - j  < 5$ )	904
Long-range ( $ i - j  \geq 5$ )	658
Total dihedral angle restraints	239
$\phi$	109
$\psi$	110
$\chi^1$	26
<b>Structure statistics</b>	
Violation occurrences	
Distance constraints ( $> 0.3 \text{ \AA}$ )	0
Dihedral angle constraints ( $> 5^\circ$ )	$0.15 \pm 0.37$
RMSD to best structure ( $\text{\AA}$ )	
All backbone atoms	$0.92 \pm 0.20$
All heavy atoms	$1.28 \pm 0.14$
Backbone atoms in secondary structures	$0.51 \pm 0.11$
Heavy atoms in secondary structures	$1.12 \pm 0.11$
Ramachandran statistics	
Residues in most favoured regions (%)	91.3
Residues in additional allowed regions (%)	8.7
Residues in generously allowed regions (%)	0
Residues in disallowed regions (%)	0

Table 1. NMR-derived restraints and structural statistics for the 20-best water-refined structures of SPAG1-TPR1<sub>opt</sub>.

	SPAG1-TPR1 <sub>opt</sub> :DASRMEEVD		SPAG1-TPR1 <sub>opt</sub> :SGPTIEEVD	
	SPAG1-TPR1 <sub>opt</sub>	DASRMEEVD	SPAG1-TPR1 <sub>opt</sub>	SGPTIEEVD
<b>Active residues</b>				
Total	9	9		9
Residues	214 217 220 221 243 254 277 305 308	all	214 217 220 221 243 254 277 305 306	all
<b>Passive residues</b>				
Total	20	0	20	0
Residues	210 213 216 225 240 241 242 246 247 250 251 275 276 278 280 281 306 307 309 310		210 213 216 225 240 241 242 246 247 250 251 275 276 278 280 281 306 307 308 309	
<b>AIR</b> (Ambiguous Intermolecular Restraints)	9	9	9	9
<b>UIR</b> (Unambiguous Intermolecular Restraints)		27	24	
Coming from intermolecular NOEs		17	14	
<b>Top-5 structures statistics</b>				
HADDOCK score	-114.959 ± 6.842		-103.447 ± 1.751	
AIR violation (> 0.3 Å)	0.4 ± 0.5		0	
UIR violation (> 0.3 Å)	0		0	
RMSD to best structure (Å)				
All backbone atoms	0.74 ± 0.15		0.67 ± 0.05	
All heavy atoms	1.07 ± 0.13		0.95 ± 0.07	

Table 2. Summary of restraints used in HADDOCK and structural statistics for the 5-best structures of SPAG1-TPR1<sub>opt</sub> in complex with the DASRMEEVD and SGPTIEEVD peptides.



## REFERENCES

- (1) Ponting, C. P.; Russell, R. R. The Natural History of Protein Domains. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31* (1), 45–71. <https://doi.org/10.1146/annurev.biophys.31.082901.134314>.
- (2) Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. Structure-Based Maximal Affinity Model Predicts Small-Molecule Druggability. *Nat. Biotechnol.* **2007**, *25* (1), 71–75. <https://doi.org/10.1038/nbt1273>.
- (3) Structural Genomics Consortium; Architecture et Fonction des Macromolécules Biologiques; Berkeley Structural Genomics Center; China Structural Genomics Consortium; Integrated Center for Structure and Function Innovation; Israel Structural Proteomics Center; Joint Center for Structural Genomics; Midwest Center for Structural Genomics; New York Structural GenomiX Research Center for Structural Genomics; Northeast Structural Genomics Consortium; Oxford Protein Production Facility; Protein Sample Production Facility, Max Delbrück Center for Molecular Medicine; RIKEN Structural Genomics/Proteomics Initiative; SPINE2-Complexes. Protein Production and Purification. *Nat. Methods* **2008**, *5* (2), 135–146. <https://doi.org/10.1038/nmeth.f.202>.
- (4) Assenberg, R.; Wan, P. T.; Geisse, S.; Mayr, L. M. Advances in Recombinant Protein Expression for Use in Pharmaceutical Research. *Curr. Opin. Struct. Biol.* **2013**, *23* (3), 393–402. <https://doi.org/10.1016/j.sbi.2013.03.008>.
- (5) Wingfield, P. T. Overview of the Purification of Recombinant Proteins. *Curr. Protoc. Protein Sci.* **2015**, *80* (1). <https://doi.org/10.1002/0471140864.ps0601s80>.
- (6) Goldenzweig, A.; Goldsmith, M.; Hill, S. E.; Gertman, O.; Laurino, P.; Ashani, Y.; Dym, O.; Unger, T.; Albeck, S.; Prilusky, J.; Lieberman, R. L.; Aharoni, A.; Silman, I.; Sussman, J. L.; Tawfik, D. S.; Fleishman, S. J. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* **2016**, *63* (2), 337–346. <https://doi.org/10.1016/j.molcel.2016.06.012>.
- (7) D'Andrea, L. D.; Regan, L. TPR Proteins: The Versatile Helix. *Trends Biochem Sci* **2003**, *28* (12), 655–662. <https://doi.org/S0968000403002731> [pii].
- (8) Das, A. K.; Cohen, P. W.; Barford, D. The Structure of the Tetratricopeptide Repeats of Protein Phosphatase 5: Implications for TPR-Mediated Protein-Protein Interactions. *EMBO J* **1998**, *17* (5), 1192–1199. <https://doi.org/10.1093/emboj/17.5.1192>.
- (9) Main, E. R. G.; Xiong, Y.; Cocco, M. J.; D'Andrea, L.; Regan, L. Design of Stable  $\alpha$ -Helical Arrays from an Idealized TPR Motif. *Structure* **2003**, *11* (5), 497–508. [https://doi.org/10.1016/S0969-2126\(03\)00076-5](https://doi.org/10.1016/S0969-2126(03)00076-5).
- (10) Kajander, T.; Cortajarena, A. L.; Mochrie, S.; Regan, L. Structure and Stability of Designed TPR Protein Superhelices: Unusual Crystal Packing and Implications for Natural TPR Proteins. *Acta Crystallogr. D Biol. Crystallogr.* **2007**, *63* (7), 800–811. <https://doi.org/10.1107/S09074444907024353>.
- (11) Chen, S.; Smith, D. F. Hop as an Adaptor in the Heat Shock Protein 70 (Hsp70) and Hsp90 Chaperone Machinery. *J Biol Chem* **1998**, *273* (52), 35194–35200.
- (12) Brinker, A.; Scheuffler, C.; Von Der Mulbe, F.; Fleckenstein, B.; Herrmann, C.; Jung, G.; Moarefi, I.; Hartl, F. U. Ligand Discrimination by TPR Domains. Relevance and Selectivity of EEVD-Recognition in Hsp70 x Hop x Hsp90 Complexes. *J Biol Chem* **2002**, *277* (22), 19265–19275. <https://doi.org/10.1074/jbc.M109002200> M109002200 [pii].
- (13) Scheuffler, C.; Brinker, A.; Bourenkov, G.; Pegoraro, S.; Moroder, L.; Bartunik, H.; Hartl, F. U.; Moarefi, I. Structure of TPR Domain-Peptide Complexes: Critical Elements in the Assembly of the Hsp70-Hsp90 Multichaperone Machine. *Cell* **2000**, *101* (2), 199–210. [https://doi.org/S0092-8674\(00\)80830-2](https://doi.org/S0092-8674(00)80830-2) [pii] 10.1016/S0092-8674(00)80830-2.
- (14) Henri, J.; Chagot, M. E.; Bourguet, M.; Abel, Y.; Terral, G.; Maurizy, C.; Aigueperse, C.; Georgescauld, F.; Vandermoere, F.; Saint-Fort, R.; Behm-Ansmant, I.; Charpentier, B.; Pradet-Balade, B.; Verheggen, C.; Bertrand, E.; Meyer, P.; Cianferani, S.; Manival, X.; Quinteret, M. Deep Structural Analysis of RPAP3 and PIH1D1, Two Components of the HSP90 Co-Chaperone R2TP Complex. *Structure* **2018**, *26* (9), 1196–1209 e8. <https://doi.org/10.1016/j.str.2018.06.002>.
- (15) Chagot, M. E.; Dos Santos Morais, R.; Dermouche, S.; Lefebvre, D.; Manival, X.; Chipot, C.; Dehez, F.; Quinteret, M. Binding Properties of the Quaternary Assembly Protein SPAG1. *Biochem J* **2019**, *476* (11), 1679–1694. <https://doi.org/10.1042/BCJ20190198>.
- (16) Back, R.; Dominguez, C.; Rothe, B.; Bobo, C.; Beauflis, C.; Morera, S.; Meyer, P.; Charpentier, B.; Branlant, C.; Allain, F. H.; Manival, X. High-Resolution Structural Analysis Shows How Tah1 Tethers Hsp90 to the R2TP Complex. *Structure* **2013**, *21* (10), 1834–1847. <https://doi.org/10.1016/j.str.2013.07.024>.

- (17) Schmid, A. B.; Lagleder, S.; Grawert, M. A.; Rohl, A.; Hagn, F.; Wandinger, S. K.; Cox, M. B.; Demmer, O.; Richter, K.; Groll, M.; Kessler, H.; Buchner, J. The Architecture of Functional Modules in the Hsp90 Co-Chaperone Sti1/Hop. *EMBO J* **2012**, *31* (6), 1506–1517. <https://doi.org/10.1038/emboj.2011.472>.
- (18) Tripathi, A.; Mandon, E. C.; Gilmore, R.; Rapoport, T. A. Two Alternative Binding Mechanisms Connect the Protein Translocation Sec71-Sec72 Complex with Heat Shock Proteins. *J Biol Chem* **2017**, *292* (19), 8007–8018. <https://doi.org/10.1074/jbc.M116.761122>.
- (19) Horani, A.; Ustione, A.; Huang, T.; Firth, A. L.; Pan, J.; Gunsten, S. P.; Haspel, J. A.; Piston, D. W.; Brody, S. L. Establishment of the Early Cilia Preassembly Protein Complex during Motile Ciliogenesis. *Proc Natl Acad Sci U S A* **2018**, *115* (6), E1221–E1228. <https://doi.org/10.1073/pnas.1715915115>.
- (20) Knowles, M. R.; Ostrowski, L. E.; Loges, N. T.; Hurd, T.; Leigh, M. W.; Huang, L.; Wolf, W. E.; Carson, J. L.; Hazucha, M. J.; Yin, W.; Davis, S. D.; Dell, S. D.; Ferkol, T. W.; Sagel, S. D.; Olivier, K. N.; Jahnke, C.; Olbrich, H.; Werner, C.; Raidt, J.; Wallmeier, J.; Pennekamp, P.; Dougherty, G. W.; Hjeij, R.; Gee, H. Y.; Otto, E. A.; Halbritter, J.; Chaki, M.; Diaz, K. A.; Braun, D. A.; Porath, J. D.; Schueler, M.; Baktai, G.; Griese, M.; Turner, E. H.; Lewis, A. P.; Bamshad, M. J.; Nickerson, D. A.; Hildebrandt, F.; Shendure, J.; Omran, H.; Zariwala, M. A. Mutations in SPAG1 Cause Primary Ciliary Dyskinesia Associated with Defective Outer and Inner Dynein Arms. *Am J Hum Genet* **2013**, *93* (4), 711–720. <https://doi.org/10.1016/j.ajhg.2013.07.025>.
- (21) Lin, W.; Zhou, X.; Zhang, M.; Li, Y.; Miao, S.; Wang, L.; Zong, S.; Koide, S. S. Expression and Function of the HSD-3.8 Gene Encoding a Testis-Specific Protein. *Mol Hum Reprod* **2001**, *7* (9), 811–818.
- (22) Maurizy, C.; Quinternet, M.; Abel, Y.; Verheggen, C.; Santo, P. E.; Bourguet, M.; C. F. Paiva A; Bragantini, B.; Chagot, M. E.; Robert, M. C.; Abeza, C.; Fabre, P.; Fort, P.; Vandermoere, F.; M. F. Sousa P; Rain, J. C.; Charpentier, B.; Cianferani, S.; Bandeiras, T. M.; Pradet-Balade, B.; Manival, X.; Bertrand, E. The RPAP3-Cterminal Domain Identifies R2TP-like Quaternary Chaperones. *Nat Commun* **2018**, *9* (1), 2093. <https://doi.org/10.1038/s41467-018-04431-1>.
- (23) Rodríguez, C. F.; Llorca, O. RPAP3 C-Terminal Domain: A Conserved Domain for the Assembly of R2TP Co-Chaperone Complexes. *Cells* **2020**, *9* (5). <https://doi.org/10.3390/cells9051139>.
- (24) Chagot, M. E.; Jacquemin, C.; Branlant, C.; Charpentier, B.; Manival, X.; Quinternet, M. (1)H, (15)N and (13)C Resonance Assignments of the Two TPR Domains from the Human RPAP3 Protein. *Biomol NMR Assign* **2015**, *9* (1), 99–102. <https://doi.org/10.1007/s12104-014-9552-4>.
- (25) Manival, X.; Jacquemin, C.; Charpentier, B.; Quinternet, M. (1)H, (15)N and (13)C Resonance Assignments of the Yeast Pih1 and Tah1 C-Terminal Domains Complex. *Biomol NMR Assign* **2015**, *9* (1), 71–73. <https://doi.org/10.1007/s12104-014-9547-1>.
- (26) Jimenez, B.; Ugwu, F.; Zhao, R.; Orti, L.; Makhnevych, T.; Pineda-Lucena, A.; Houry, W. A. Structure of Minimal Tetratricopeptide Repeat Domain Protein Tah1 Reveals Mechanism of Its Interaction with Pih1 and Hsp90. *J Biol Chem* **2012**, *287* (8), 5698–5709. <https://doi.org/10.1074/jbc.M111.287458>.
- (27) Pal, M.; Morgan, M.; Phelps, S. E.; Roe, S. M.; Parry-Morris, S.; Downs, J. A.; Polier, S.; Pearl, L. H.; Prodromou, C. Structural Basis for Phosphorylation-Dependent Recruitment of Tel2 to Hsp90 by Pih1. *Structure* **2014**. <https://doi.org/10.1016/j.str.2014.04.001>.
- (28) Fabczak, H.; Osinka, A. Role of the Novel Hsp90 Co-Chaperones in Dynein Arms' Preassembly. *Int. J. Mol. Sci.* **2019**, *20* (24). <https://doi.org/10.3390/ijms20246174>.
- (29) Boulon, S.; Pradet-Balade, B.; Verheggen, C.; Molle, D.; Boireau, S.; Georgieva, M.; Azzag, K.; Robert, M. C.; Ahmad, Y.; Neel, H.; Lamond, A. I.; Bertrand, E. HSP90 and Its R2TP/Prefoldin-like Cochaperone Are Involved in the Cytoplasmic Assembly of RNA Polymerase II. *Mol Cell* **2010**, *39* (6), 912–924. <https://doi.org/10.1016/j.molcel.2010.08.023>.
- (30) Benbahouche Nel, H.; Iliopoulos, I.; Torok, I.; Marhold, J.; Henri, J.; Kajava, A. V.; Farkas, R.; Kempf, T.; Schnolzer, M.; Meyer, P.; Kiss, I.; Bertrand, E.; Mechler, B. M.; Pradet-Balade, B. Drosophila Spag Is the Homolog of RNA Polymerase II-Associated Protein 3 (RPAP3) and Recruits the Heat Shock Proteins 70 and 90 (Hsp70 and Hsp90) during the Assembly of Cellular Machineries. *J Biol Chem* **2014**, *289* (9), 6236–6247. <https://doi.org/10.1074/jbc.M113.499608>.
- (31) Martino, F.; Pal, M.; Munoz-Hernandez, H.; Rodriguez, C. F.; Nunez-Ramirez, R.; Gil-Carton, D.; Degliesposti, G.; Skehel, J. M.; Roe, S. M.; Prodromou, C.; Pearl, L. H.; Llorca, O. RPAP3 Provides a Flexible Scaffold for Coupling HSP90 to the Human R2TP Co-Chaperone Complex. *Nat Commun* **2018**, *9* (1), 1501. <https://doi.org/10.1038/s41467-018-03942-1>.
- (32) Muñoz-Hernández, H.; Pal, M.; Rodríguez, C. F.; Fernández-Leiro, R.; Prodromou, C.; Pearl, L. H.; Llorca, O. Structural Mechanism for Regulation of the AAA-ATPases RUVBL1-RUVBL2 in the R2TP Co-Chaperone Revealed by Cryo-EM. *Sci. Adv.* **2019**, *5* (5), eaaw1616. <https://doi.org/10.1126/sciadv.aaw1616>.

- (33) Yamaguchi, H.; Oda, T.; Kikkawa, M.; Takeda, H. Systematic Studies of All PIH Proteins in Zebrafish Reveal Their Distinct Roles in Axonemal Dynein Assembly. *Elife* **2018**, *7*.  
<https://doi.org/10.7554/eLife.36979>.
- (34) Horejsi, Z.; Stach, L.; Flower, T. G.; Joshi, D.; Flynn, H.; Skehel, J. M.; O'Reilly, N. J.; Ogrodowicz, R. W.; Smerdon, S. J.; Boulton, S. J. Phosphorylation-Dependent PIH1D1 Interactions Define Substrate Specificity of the R2TP Cochaperone Complex. *Cell Rep* **2014**, *7* (1), 19–26.  
<https://doi.org/10.1016/j.celrep.2014.03.013>.
- (35) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat Methods* **2015**, *12* (1), 7–8. <https://doi.org/10.1038/nmeth.3213>.
- (36) Diebold, M. L.; Fribourg, S.; Koch, M.; Metzger, T.; Romier, C. Deciphering Correct Strategies for Multiprotein Complex Assembly by Co-Expression: Application to Complexes as Large as the Histone Octamer. *J Struct Biol* **2011**. [https://doi.org/S1047-8477\(11\)00040-2](https://doi.org/S1047-8477(11)00040-2) [pii] 10.1016/j.jsb.2011.02.001.
- (37) Berjanskii, M. V.; Neal, S.; Wishart, D. S. PREDITOR: A Web Server for Predicting Protein Torsion Angle Restraints. *Nucleic Acids Res* **2006**, *34* (Web Server issue), W63–9. <https://doi.org/10.1093/nar/gkl341>.
- (38) Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *J Biomol NMR* **2013**, *56* (3), 227–241. <https://doi.org/10.1007/s10858-013-9741-y>.
- (39) Guntert, P. Automated NMR Structure Calculation with CYANA. *Methods Mol Biol* **2004**, *278*, 353–378. <https://doi.org/1-59259-809-9:353> [pii] 10.1385/1-59259-809-9:353.
- (40) Bertini, I.; Case, D. A.; Ferella, L.; Giachetti, A.; Rosato, A. A Grid-Enabled Web Portal for NMR Structure Refinement with AMBER. *Bioinformatics* **2011**, *27* (17), 2384–2390.  
<https://doi.org/10.1093/bioinformatics/btr415>.
- (41) de Vries, S. J.; van Dijk, A. D.; Krzeminski, M.; van Dijk, M.; Thureau, A.; Hsu, V.; Wassenaar, T.; Bonvin, A. M. HADDOCK versus HADDOCK: New Features and Performance of HADDOCK2.0 on the CAPRI Targets. *Proteins* **2007**, *69* (4), 726–733. <https://doi.org/10.1002/prot.21723>.
- (42) Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8, 2015.
- (43) Rodrigues, J. P.; Trellet, M.; Schmitz, C.; Kastiris, P.; Karaca, E.; Melquiond, A. S.; Bonvin, A. M. Clustering Biomolecular Complexes by Residue Contacts Similarity. *Proteins* **2012**, *80* (7), 1810–1817.  
<https://doi.org/10.1002/prot.24078>.
- (44) Kandoi, G.; Acencio, M. L.; Lemke, N. Prediction of Druggable Proteins Using Machine Learning and Systems Biology: A Mini-Review. *Front. Physiol.* **2015**, *6*. <https://doi.org/10.3389/fphys.2015.00366>.
- (45) Darby, J. F. Solution Structure of the Hop TPR2A Domain and Investigation of Target Druggability by NMR, Biochemical and in Silico Approaches. *Sci. Rep.* **2020**, *13*.
- (46) Kawachi, T.; Tanaka, S.; Fukuda, A.; Sumii, Y.; Setiawan, A.; Kotoku, N.; Kobayashi, M.; Arai, M. Target Identification of the Marine Natural Products Dictyoceratin-A and -C as Selective Growth Inhibitors in Cancer Cells Adapted to Hypoxic Environments. *Mar. Drugs* **2019**, *17* (3), 163.  
<https://doi.org/10.3390/md17030163>.