Check for updates

FEBS Letters

FEBS PRESS science publishing by scientists

Experimentally based structural model of Yih1 provides insight into its function in controlling the key translational regulator Gcn2

Elena Harjes[1,2], Geoffrey B. Jameson[1,2], Yi-Hsuan Tu[1], Natalie Burr[1], Trevor S. Loo[1], Alexander K. Goroncy[1,*], Patrick J.B. Edwards[1], Stefan Harjes[1], Ben Munro[1], Christoph Göbl[3], Evelyn Sattlegger[2,4] and Gillian E. Norris[1,2]

1 School of Fundamental Sciences, Massey University, Palmerston North, New Zealand
2 Maurice Wilkins Centre for Molecular BioDiscovery, Massey University, Palmerston North, New Zealand
3 Centre for Free Radical Research, Department of Pathology, University of Otago, Christchurch, New Zealand
4 School of Natural and Computational Sciences, Massey University, Auckland, New Zealand

Correspondence
E. Harjes, Gillian E Norris and Geoffrey B Jameson, School of Fundamental Sciences, Massey University, Private Bag 11 222, Palmerston North, New Zealand
Tel: +64 6 356 9099 (EH); +64 (06) 356 9099 (GEN, GBJ)
E-mails: e.harjes@massey.ac.nz (EH); g.norris@massey.ac.nz (GEN); g.b.jameson@massey.ac.nz (GBJ)

Present address
* Chemistry Department, University of Wyoming, 1000 E. University Avenue, Laramie, WY, 82071, USA

(Received 18 June 2020, revised 21 October 2020, accepted 2 November 2020, available online 22 November 2020)

doi:10.1002/1873-3468.13990

Edited by Christian Griesinger

Yeast impact homolog 1 (Yih1), or IMPACT in mammals, is part of a conserved regulatory module controlling the activity of General Control Nonderepressible 2 (Gcn2), a protein kinase that regulates protein synthesis. Yih1/IMPACT is implicated not only in many essential cellular processes, such as neuronal development, immune system regulation and the cell cycle, but also in cancer. Gcn2 must bind to Gcn1 in order to impair the initiation of protein translation. Yih1 hinders this key Gcn1-Gcn2 interaction by binding to Gcn1, thus preventing Gcn2-mediated inhibition of protein synthesis. Here, we solved the structures of the two domains of *Saccharomyces cerevisiae* Yih1 separately using Nuclear Magnetic Resonance and determined the relative positions of the two domains using a range of biophysical methods. Our findings support a compact structural model of Yih1 in which the residues required for Gcn1 binding are buried in the interface. This model strongly implies that Yih1 undergoes a large conformational rearrangement from a latent closed state to a primed open state to bind Gcn1. Our study provides structural insight into the interactions of Yih1 with partner molecules.

Keywords: mass spectrometry; nuclear magnetic resonance; protein conformation; protein structure; small-angle X-ray scattering; structural biology; structure; function; translation initiation; translation regulation; Yeast IMPACT Homolog 1; Yih1

Yeast IMPACT Homolog 1, or Yih1, is the ortholog of the mammalian protein IMPACT (imprinted gene with Ancient Domain). Yih1/IMPACT is a negative regulator of General Control Nonderepressible 2 (Gcn2), a protein involved in reducing initiation of translation during times of amino acid starvation [1,2]. The Yih1/IMPACT-Gcn2 module has recently attracted interest due to its implication in central biological functions, such as neuronal development [3] and the immune system [1]. So far, Yih1 has been

Abbreviations
ADH, adipic acid dihydrazide; DMTMM, 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride; Gcn2, General Control Nonderepressible 2; NMR, Nuclear Magnetic Resonance; NOE, Nuclear Overhauser Effect; PREs, paramagnetic relaxation enhancements; RDCs, residual dipolar couplings; SAXS, Small Angle X-ray Scattering; TFA, trifluoroacetic acid; Yih1, Yeast impact homolog 1.

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

studied in more detail than IMPACT. When Yih1 expression is induced in yeast, it binds to Gcn1 [4], a protein that is absolutely necessary for promoting Gcn2 activation [1]. While studies suggest that this interaction of Gcn1 with Gcn2 is necessary for Gcn2 to receive the starvation signal, the exact mechanism of Gcn2 activation is not yet known [5,6]. Upon signal recognition, Gcn2 becomes activated and phosphorylates the eukaryotic translation initiation factor 2α (eIF2α), consequently reducing global protein synthesis [1]. Translation of specific mRNAs that code for transcription regulators is concomitantly enhanced, which then changes the transcription profile in the cell. When Yih1 binds to Gcn1, it prevents Gcn1-Gcn2 interaction, and consequently Gcn2 is unable to be activated, as illustrated in Fig. 1A,C. Prevention of eIF2α phosphorylation allows protein synthesis to proceed at the maximum rate.

Besides interacting with Gcn1, Yih1 also is reported to bind to monomeric actin (G-actin) *in vivo* [4]. Studies suggest that when Yih1 is released from G-actin (by an as yet unknown mechanism), it is able to compete with Gcn2 to bind to Gcn1, thereby preventing Gcn2 activation [7]. Yih1 provides the possible link between the cytoskeleton and protein synthesis in cells by interacting with both actin and Gcn1. The actin cytoskeleton, a polymer of actin monomers undergoes constant assembly and disassembly, allowing it to respond quickly to external and internal cues (e.g., environmental stress, amino acid starvation) [8]. Because the concentrations of free Yih1 in the cell are controlled by the ability of unpolymerized actin to reversibly bind Yih1, the level of Gcn2 activity, and hence the rate of protein synthesis is also controlled by the state of the actin cytoskeleton [9]. As Yih1 binds to the ribosome, where Gcn1 and Gcn2 also appear to reside, Yih1 is ideally positioned to instantly control Gcn2 activation in the cell [10]. It has also been shown that the cell cycle is delayed at the G2/M phase in cells lacking Yih1, another cellular function of Yih1 that does not involve Gcn2 [9]. The same phenomenon was seen when a mutated form of Yih1 (E87A:D90A) was found to have enhanced affinity to the cell cycle protein cyclin-dependent kinase Cdc28 [9]. Thus, through its interaction with Gcn1, Yih1 is able to control the many biological functions of Gcn2 and, as a result, regulates multiple cellular functions. Although these have yet to be clearly defined, regions of the Yih1 sequence that have been shown to be involved in protein-protein interactions by co-precipitation assays overlap, suggesting that Yih1 may only be able to bind only one physiological partner at a time [1].

Yih1 is a two-domain protein, consisting of 258 amino acids. The N-terminal domain (residues 1–112)

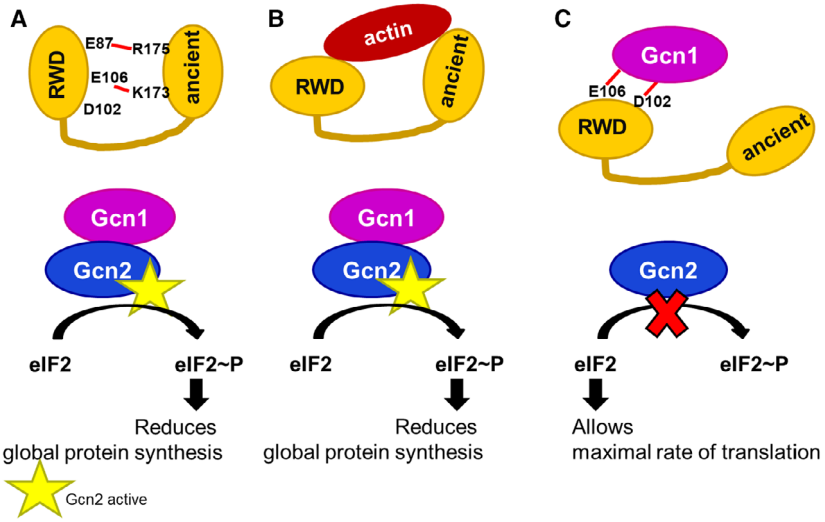


Fig. 1. Model for Yih1 interdomain movement and its consequence on Gcn2 activity. (A) Interdomain interactions within Yih1 renders it inactive, allowing cells to fully activate Gcn2. (B) Loss of Yih1 interdomain interactions allows actin to bind to Yih1. Yih1 sequestration prevents it from disrupting the Gcn1-Gcn2 interaction, thereby allowing cells to fully activate Gcn2. Activated Gcn2 phosphorylates eIF2, which leads to reduced global protein synthesis. (C) Loss of Yih1 interdomain interactions allows Gcn1 to bind to Yih1. As a consequence, Gcn2 is displaced from Gcn1 and cannot be activated, thereby allowing cells to synthesize proteins at maximal rate. It is possible that actin and Gcn1 disrupt the Yih1 interdomain interactions to allow them to bind to Yih1. *In vitro* studies indicate that an additional party may be needed for Yih1-actin interactions *in vivo*.

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

is predicted to be an RWD [11] domain, which is involved in the interaction with Gcn1 [4,7,12], although the sequence identity to known RWD structures is low (10–22%). The name RWD arose from the observations that this domain is present in many ring-finger proteins, WD-repeat proteins, and DEXDc-helicase proteins. The structure of the N-terminal RWD domain (formerly known as the GI domain) [12] of Gcn2 has been recently solved [13] and has been shown to interact with Gcn1 [5,12]. The Yih1 C-terminal domain has homology to the N-terminal half of the YigZ family of prokaryotic proteins (23–35% identity). Although the C-terminal half of YigZ contains an ACT domain, which is widely associated with the control of amino acid and nucleobase biosynthesis [14], YigZ proteins as a group have no known function. However, because the N-terminal domain of YigZ is found in many proteins from all domains of life [1], it is usually referred to as the Ancient Domain [15].

In Yih1, the Ancient Domain is linked to the RWD by a 26 amino acid linker (residues 113–138). Although the RWD domain and part of the Ancient Domain appear to be involved in the interaction of Yih1 with Gcn1 or the ribosome, only the Ancient Domain is involved in actin- and Cdc28-Yih1 interaction [1,7,9]. Although structures of each domain had been modeled based on the structures of homologous proteins [7], the structure of the whole protein has remained an enigma. Given low sequence identity (*vide supra*) to these proteins, the experimental validation of domain structures and the structure of the whole molecule are required. The stronger binding of Gcn1 to a Yih1 fragment lacking amino acids 1–67 (N-terminal) and/or 172–258, or 133–258 (C-terminal) in comparison with the full-length protein, indicates that binding may be somewhat hindered by interdomain interactions within Yih1 [7]. Mutational analysis has shown that either Asp102 or Glu106 or both residues are required for Gcn1 interaction with Yih1 [7]. Furthermore, the simultaneous mutation of Glu87 and Asp90 to Ala increases the association of Yih1 not only with Gcn1 but with all interaction partners tested so far [7,9,10], supporting the idea that a change in protein conformation may be involved. Our results in this work provide compelling evidence for a 'closed' conformation of Yih1, mediated in part by Asp and Glu side chains, and that amino acid substitutions weakening this association of domains enhance interaction of Yih1 with Gcn1.

A range of biophysical methods [Nuclear Magnetic Resonance (NMR), Small Angle X-ray Scattering (SAXS), Cross-linking with Mass Spectrometry] was used to elucidate the structural model of Yih1 as a somewhat compact two-domain protein. Incorporating knowledge from site-directed mutagenesis and binding studies has allowed us to propose a functional model in which conformational rearrangements are required to convert latent Yih1 into a protein primed for Gcn1 binding. Yih1 binding to Gcn1 in turn prevents binding to Gcn2, which hampers phosphorylation of eIF2α (Fig. 1), thereby providing a mechanism for controlling protein synthesis *via* Gcn2.

Materials and methods

Protein cloning, expression and purification

The sequence of Yih1 used in this work is UniProtKB—P25637. Wild-type Yih1 gene was cloned into the pET28a vector. All amino acid substitutions were constructed using the QuikChange protocol (Stratagene, Bellingham, WA, USA) and verified by DNA sequencing. A full description of protein expression and purification, including ^{15}N- and ^{13}C-labeled protein is provided in Supporting Information and prior publications [16].

NMR spectroscopy, SAXS and structural calculation

Standard 3D heteronuclear NMR experiments were used for chemical shifts assignment. 3D ^{15}N-edited (with 80 and 200 ms mixing times) and ^{13}C-edited [17] (100 and 200 ms mixing times), as well 2D-NOESY experiments, were used to collect distance restraints. All experiments were performed at 25 °C on 700-MHz Bruker Avance NMR spectrometer with a 3-channel inverse-detect cryoprobe. Phage (ASLA Biotech Ltd., Riga, Latvia) were used to create a partially aligned medium enabling the collection of residual dipolar couplings (RDCs) using IPAP-HSQC experiments [18]. The initial alignment tensor components were extracted with PALES [19] and subsequently a grid search was performed with CNS as described [20].

Initially structures of each of the two domains were determined separately using CNS [21] with manually assigned Nuclear Overhauser Effect (NOE)s and RDC restraints. Additional NOEs were found using ATNOS-Candid [22,23] software from UNIO package as described [16] for two domains together. SAXS was measured on the Australian Synchrotron SAXS/WAXS beam line, using the same conditions as previously described [24]. Scattering data were collected from Yih1, ~ 3.2 mg mL^{-1}, equilibrated in 10 mM KH_2PO_4/K_2HPO_4, pH 7.2, 20 mM NaCl, 1 mM EDTA, 1 mM DTT after SEC (Superdex 200 5/150; GE Healthcare, Chicago, IL, USA), and the scattered intensity (I) was plotted against the scattering vector q ($q = 4\pi\ \lambda^{-1}$ sin θ). All NOEs, RDCs and SAXS data were

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

then combined to calculate structures using XPLOR-NIH [25] with the SAXS module [26].

^{1}H-^{15}N T_1 relaxation experiments were conducted to measure paramagnetic relaxation enhancements (PREs) as a function of concentration (1–15 mM) of the paramagnetic agent Gd(DTPA-BMA), gadolinium(III) diethylenetriaminepentaacetic acid bismethylamide, in aqueous solution [27]. The Gd(DTPA-BMA) was obtained from Omniscan (GE Healthcare) and purified using the published protocol [28]. Assignment was done using CARA [29] and relaxation times were analyzed using CCPN [30]. The T_1 dependence on concentration was analyzed in Excel. All NMR spectra were collected in the following buffer: 10 mM KH_2PO_4/K_2HPO_4, pH 7.2, 20 mM NaCl, 1 mM EDTA, 1 mM DTT. PREs were back-calculated from the structures as described [31].

The SSM alignment protocol [32], embedded in COOT [33], was used for structural alignments. More details about NMR and SAXS methodologies are in the Supporting Information.

Intramolecular cross-linking

Initially the amine cross-linker bis(sulfosuccinimidyl) suberate (BS^3) with a spacer arm of 11.4 Å (Thermo Fisher Scientific, Waltham, MA, USA) was trialled using the manufacturer's instructions. Despite a number of trials, very low concentrations of monomers were seen in the products. Examination of these using mass spectrometry showed there were no intramolecular cross-links (Results not shown). A chemical cross-linking agent specific for carboxyl groups [34] was therefore trialled. Yih1 was diluted to a concentration of 250 μM (7.5 mg·mL^{-1}) in 50 mM HEPES buffer pH 7.0. Cross-linking was initiated by adding a mixture of adipic acid dihydrazide (ADH) and the activating reagent 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) at a molar ratio of 1 : 1.5 ADH : DMTMM) to the protein solution at a molar ratio of 1 : 3 Yih1 : ADH in a total volume of 100 μL. Samples were incubated at 37 °C for 1 h in a ThermoShaker at 750 r.p.m. The reaction was stopped by subjecting the samples to RP-HPLC (Jupiter C_{18} column, 4.6 mm × 250 mm with a 5 μM particle size (Phenomenex, Torrance, CA USA), using a linear water/0.1% trifluoroacetic acid (TFA)–acetonitrile/0.8% TFA gradient of 25 mL at 1 mL·min^{-1} on a Dionex Summit HPLC system (Thermo Fisher). The eluant was continuously monitored by absorbance at 280 nm and peak fractions collected by hand and analyzed by SDS/PAGE to detect fractions containing intermolecular cross-links. Those containing only monomers were analyzed by mass spectrometry (Table S1) using standard proteomic techniques to determine the number of cross-links (Table S2,S3). Data were stored in profile mode, and the total ion chromatograms were examined. Peaks within the range of 600–3000 m/z were deconvoluted using either the pMod algorithm (AGILENT MASSHUNTER Workstation Qualitative Analysis software version B.03.01; Agilent Technologies, Santa Clara, CA, USA) or Excalibur Qual Browser (Thermo Fisher Scientific). Those containing intramolecular cross-links, along with control samples that had not been exposed to cross-linking reagents, were further analyzed using standard proteomic techniques in an attempt to determine the likely location of the cross-links (PD2.1 (Thermo Fisher Scientific), Tables S2–S4).

To produce a greater quantity of cross-linked Yih1 for analysis, cross-linking was scaled and the resulting cross-linked peptides fractionated and analyzed by mass spectrometry as described in the Supporting Information.

Results

Recombinant Yih1 production

Anion-exchange chromatography of the *E. coli* cell lysate containing Yih1 resulted in a split peak containing two forms of Yih1 labeled Yih1a and Yih1b (Fig. S1); these were combined, subjected to SEC (Fig. S2), and then analyzed by SDS/PAGE (Fig. S3). Mass spectrometry showed that the peak in SEC had two main components, one with a mass of 29 018 amu and the other a mass of 29 046 amu, that is, a molecular weight difference of 28 amu (Fig. S4). Fragmentation of tryptic/chymotryptic peptides from these peaks showed that the mass difference corresponds to dimethylation of Arg144 in the Ancient Domain (fragment ^{136}TASDPITD**R**) or Lys36 (fragment ^{26}KQEDGSIIVV**K**) in the RWD domain (Figs S5,S6). Mass spectrometry showed that at both sites, dimethylation is partial and no evidence of a doubly dimethylated protein was observed. Such a modification could result in the dimethylated protein being retained on an anion-exchange column longer than the unmodified protein but as the peaks were poorly resolved, both species were observed in each pooled fraction. Size-exclusion chromatography of both pooled fractions resulted in a protein that was ~ 99% pure. Both fractions were used for the following studies.

NMR structures of RWD and Ancient domains

The assignment was done on the full-length protein with standard triple resonance NMR spectra and 95.99% of backbone atoms were assigned (full information in Table S4). All resolved peaks in the $^{1}H^{15}N$-HSQC spectrum were assigned with the exception of the spectral region corresponding to the Asn/Gln side chains. The HN signals of Ala71 (RWD), Glu114 (linker), Asp197, and Phe230. Gly231 and residues 233–

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

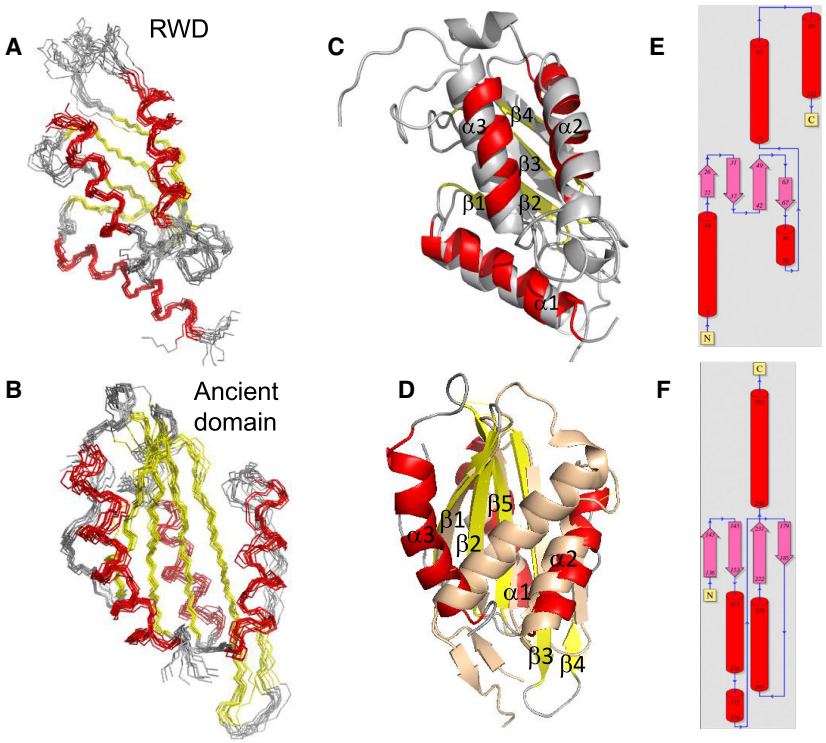


Fig. 2. Ensemble of structures of the N-terminal RWD (A) and of the C-terminal Ancient Domain (B) of Yih1 protein. α-Helices are colored red and β-strands yellow. (C) Superposition of the RWD domain of recombinant Yih1 from *S. cerevisiae* (this study) colored as in A with the structure of the RWD domain of mouse GCN2 colored gray (PDB-ID: 1ukx; 18.6% identity, RMSD 2.4 Å for superposition of 97 residues out of 114 (Yih1)) [13]. (D) Superposition of the Ancient Domain of recombinant Yih1 from *S. cerevisiae* (this study) colored as in B with conserved hypothetical protein TT1547 from *T. thermophilus* HB8 colored peach (PDB-ID: 2cve; 35.7% identity, RMSD 2.7 Å for superposition of 112 residues out of 122(Yih1)). Note that TT1547 has short N-terminal and C-terminal β strands not shared with Yih1 and a C-terminal helix α3 positioned differently to Yih1. (E), (F) Topology diagrams of, respectively, the RWD and Ancient Domain of Yih1, highlighting the distinctly different topologies. Note: the short strand β4 of the Ancient Domain of Yih1 is not captured by PDBSum. Diagrams produced by PDBSum.

239 (Ancient Domain) were missing, most likely due to chemical exchange.

As initial evaluation of NOE NMR data of the full-length Yih1 did not reveal any short-range interactions between the RWD and the Ancient Domain of Yih1, we decided to determine the structures of the two domains separately in the context of the full-length protein. The final structures of the N-terminal RWD domain (residues 1–112) and the predicted Ancient Domain (residues 137–258), calculated for a single chain using both NOE and RDC data, resembled the topology predicted by the homology models [7]. Accordingly, the RWD consists of three helices and four beta-strands forming an antiparallel β sheet (Fig. 2A), which is similar to the structure of the RWD domain of mouse Gcn2 (PDB-ID: 1ukx [13] Fig. 2C). Despite the low sequence identity (24% for 76 amino acids), the RWD domains of Yih1 and of mouse GCN2 superimpose well (using the software YASARA [35], RMSD = 1.99 Å over those 76 amino acids, involved mainly in the secondary structure). The main differences in the structure are seen in the loops, which have different lengths in the two proteins. This is particularly noticeable in the longer loop between β4 and α2, which has some helical tendency in the RWD of Yih1. There are also differences in the length of some of the other loops between RWD domains of Yih1 and of Gcn2.

The Ancient Domain consists of three α helices and five β strands, forming an antiparallel β sheet (Fig. 2B), which is similar to the structure of the Ancient Domain of the conserved hypothetical protein TT1547 from *Thermus thermophilus* HB8 (PDB-ID: 2cve Fig. 2D; RMSD of 1.72 Å over 77 residues, YASARA [35]). The topologies of the RWD and Ancient domains are, however, distinctly different (Fig. 2E,F). In other molecules, the Ancient Domain is generally fused to a C-terminal domain that is

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

topologically identical to an ACT domain [14], a domain widely involved in regulation of amino acid and nucleobase biosynthesis. Intradomain NOE data, showed that in Yih1 residues 127–136, initially assigned to the linker region by comparison with the structure of YigZ (PDB-ID: 1vi7) [36], had interactions with some residues in the Ancient Domain (Fig. S7). As a result, these amino acids were assigned to the Ancient Domain, which now comprises residues 127–258. Indeed, the difference in the position of helix 3 of the Ancient Domain in comparison with its orientation in YigZ [36] and the hypothetical protein TT1547 from *T. thermophilus* HB8 (PDB-ID: 2cve, Fig. 2D) may be a result of these interactions. The other predicted linker residues 113–126, appeared to remain flexible. There was no NMR signal for residue 114 in this region, and residues 121–124 were deemed to be flexible by Talos^{+} (http://spin.niddk.nih.gov/bax/software/TALOS). Based on the random coil index, (differences between the actual chemical shifts and those for a random coil), TALOS predicts order parameters calculated for each residue, and those residues with order parameters < 0.6 are categorized as being flexible [37].

The interaction between the two domains

The absence of interdomain NOEs would suggest there are minimal interactions between the two domains as they would independently tumble in solution, and conformational motions of the nonrigid interface would tend to extinguish these NOEs. We turned initially to SAXS to provide information about the overall shape of the molecule.

The Guinier plot (Fig. 3A) revealed a homogeneous protein solution and a radius of gyration of 20.96 Å, a value too small to result from two independently moving domains of similar size; moreover, the Kratky plot was inconsistent with two independently moving domains but consistent with a moderately compact structure (Fig. 3B). The fact that the curve does not approach zero at high q-values indicates also some flexibility in the two domains interface or subdomain dynamics within the individual domains. The pair distribution plot ($P(r)$ versus r) revealed a maximum dimension of 63 Å, and an asymmetric distribution of distances (Fig. 3C). Finally, using the NMR-determined structures of the two domains, one domain was moved around the other using the MASSHA version 2.3 [38] tool in ATSAS, and the calculated SAXS profile compared with the observed SAXS data. The SAXS data could be fitted to a relatively compact model with a maximum dimension of 62.4 Å (Fig. 3D) and are inconsistent with any model corresponding to two domains tumbling independently in the solution. Furthermore, any loosening of the interactions between the two domains, introduced greater deviations in the calculated SAXS pattern compared to the experimental profile.

Mutational analysis provided additional evidence for interaction between the two domains (Fig. 4 and Figs S8,S9), information that was not available at the time of modeling the SAXS data. Mutation of Gly96 in the RWD domain to either Ala or Ser affected the chemical shifts (^{1}H-^{15}N) of some residues in the Ancient Domain (Figs 4 and Fig S8A). In particular, both mutations attenuate the signal of Gly190 and, as expected, cause loss of the signal for Gly96. Additionally, the mutation G96S resulted in alternative conformations of the HN atoms of Trp135 and Gly134. Overall, mutating Gly96 (in the RWD domain) affects the chemical environment of some residues in the Ancient Domain. On the other hand, mutating Ile127 and Ile235 (in the Ancient Domain) to Ala influenced chemical shifts of several residues in the RWD domain, in particular, those of residues in helices two and three (Fig. S8B,C). A full tabulation of chemical shift perturbation data for the G96A, G96S, I127A, and I235A mutations is given in Table S6 and Fig. S9.

In addition, many attempts to express the protein as two separate domains using different truncations of the flexible linker failed to yield soluble folded protein. Together these results strongly suggest that the two domains are interacting with each other in wild-type Yih1, and mutation of Gly96, which lies in Ramachandran territory forbidden to the more conformationally restricted Ala and Ser, may change the interdomain interface. These results, however, do not establish which parts of the two domains are interacting.

To confirm that there was interaction between the domains, paramagnetic solvent relaxation enhancement [27] experiments were carried out. Addition of the paramagnetic agent Gd(DTPA-BMA) to the solvent increases the relaxation times of solvent-accessible residues, whereas nonaccessible residues are much less affected [39]. Surface residues of the individual domains that show only small relaxation time enhancement in paramagnetic solvents are therefore likely to be involved in the interdomain interface. At this point, we were able to confidently identify only one residue meeting these criteria. The surface residue, Gly96, in the RWD domain, had a relaxation time that hardly changed in the paramagnetic solvent (Fig. 5). It is situated in a tight loop between helices two and three of the RWD domain (Fig. S10), the helices that are affected by the mutations made to residues in the Ancient Domain described above. That Gly96 is not

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

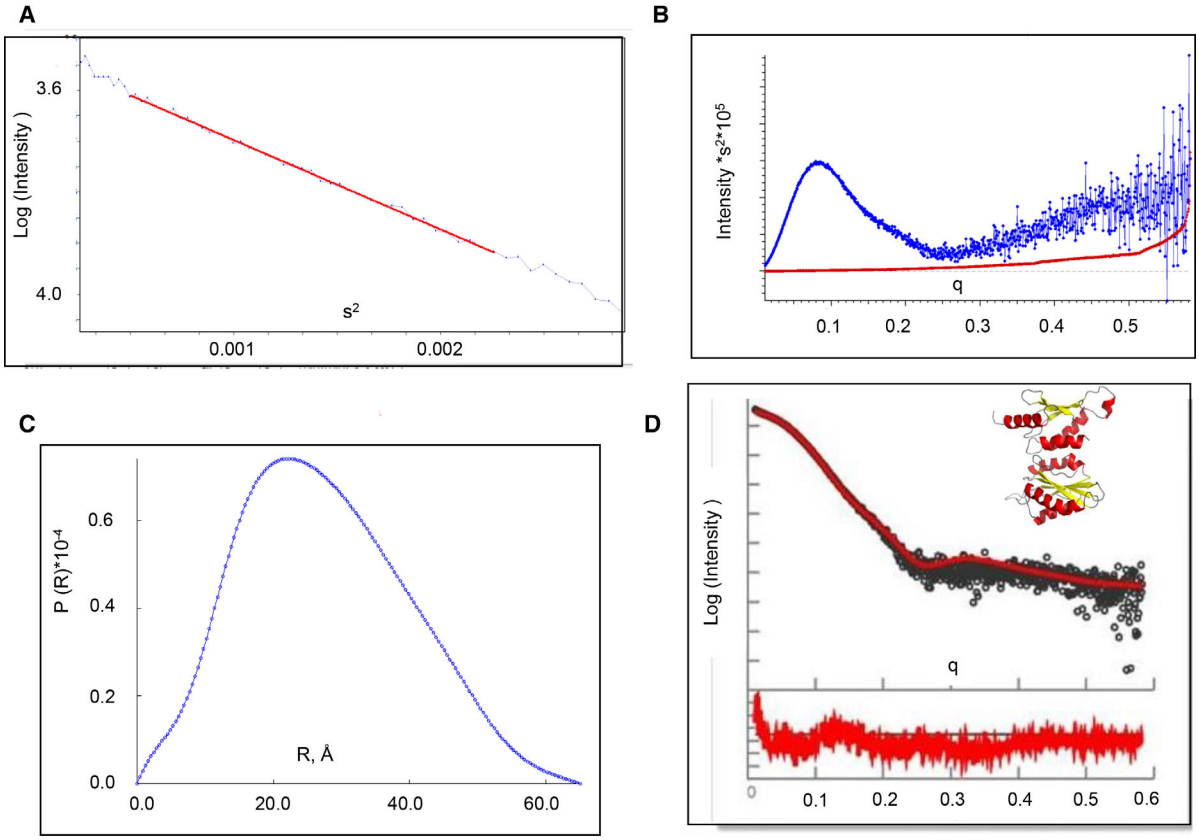


Fig. 3. SAXS data and model for Yih1. (A) Guinier plot showing absence of aggregation. (B) Kratky plot showing domains are not undergoing uncorrelated motions. (C) Pair distribution plot, showing molecular asymmetry. (D) Fit of SAXS data to the first generation model for the positioning of the two domains of Yih1 [56] (χ^2 is 2.75). The inset shows the relative alignment of the two domains whose structures were determined by NMR methods.

affected by the paramagnetic solvent strongly suggests it may be part of the domain interface. Subsequently, one possible weak NOE between the H(N) of Leu100 (in the RWD domain) and the H(N) of Trp135 (in the Ancient Domain) was identified (Fig. S11).

Structural model of the two-domain protein

Intradomain NOEs, SAXS data [26], RDC data and the one possible interdomain NOE were included in a structural calculation using XPLOR-NIH [25]. As RDCs depend on the angle between the magnetic field and the H-N bond direction, the fit of all the measured RDCs to all possible orientations between the two domains was calculated, resulting in 250 structures. Eleven were of low energy and could be grouped into two clusters, both of which had Gly96 situated in the interdomain interface (Fig. S12). However, only one of these clusters, comprised of three structures, was consistent with the solvent accessibility data (Fig. S12A). Plots of calculated versus observed RDCs are provided in Fig. S13, where the correlation coefficient for the accepted ensemble is 99.5%. The RDC analysis provided additional evidence for the limited flexibility between two domains as the alignment tensor components were very similar in the initial PALES [19] analysis and the same values of the alignment tensor's components were used in structural calculations for both domains. The values of the alignment tensor components (Da, Dr), calculated using the bestFit algorithm from PALES are very similar for each separate domains (RWD: -1.1×10^{-3}, -4.4×10^{-4}, Q-factor: 0.003); (Ancient Domain: -1.1×10^{-3}, -4.2×10^{-4}, Q-factor 0.015), and for the final full-length ensemble: -1.2×10^{-3}, -5.8×10^{-4}, Q-factor: 0.013). Furthermore, we did not observe any difference in the relaxation rates R_1 in the RWD and Ancient Domain (Fig. S13B). In contrast, had the two domains been moving independently we would have observed significantly different degrees of alignment and relaxation rates [40].

Nuclear Magnetic Resonance signals that disappeared as the concentrations of paramagnetic molecules

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

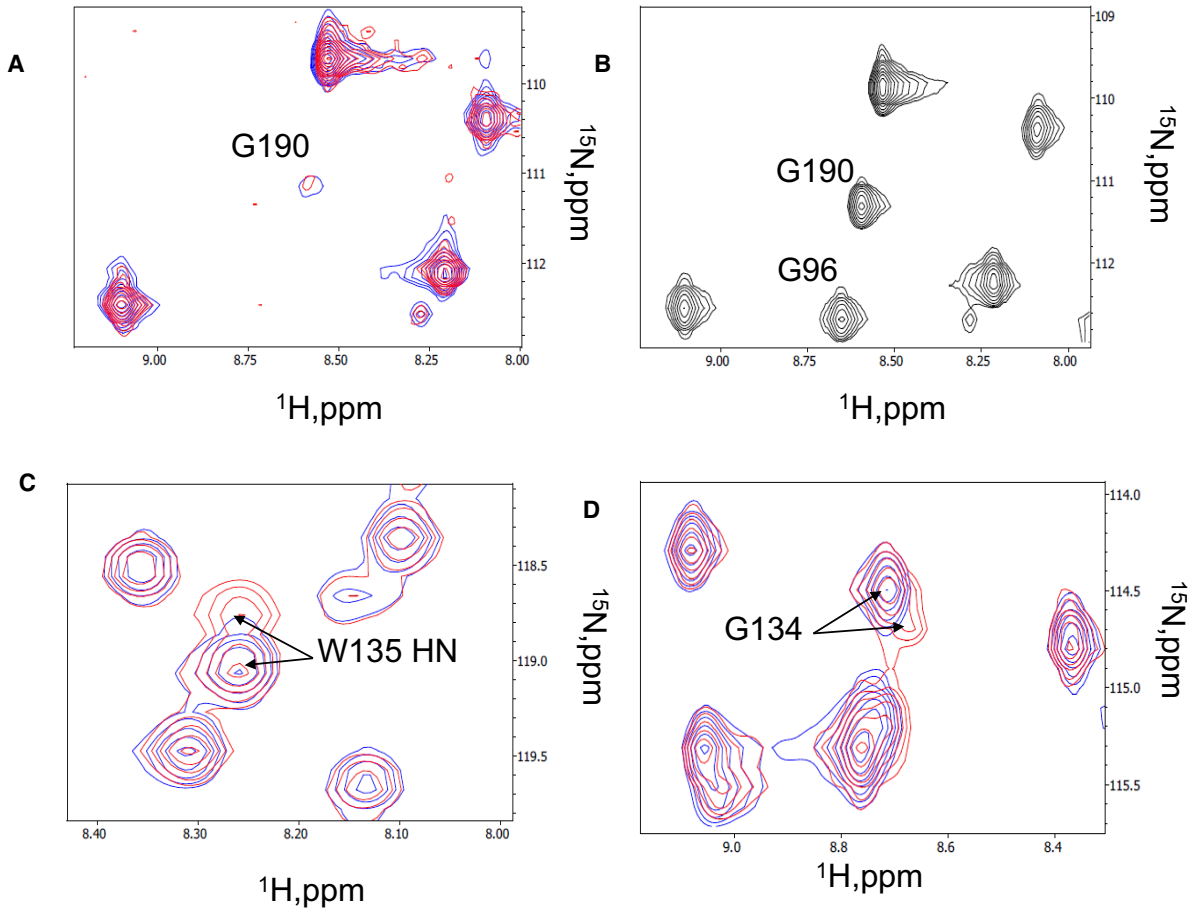


Fig. 4. Mutations in the RWD domain affect chemical shifts of the Ancient Domain for Yih1. Expanded area of the NMR ^{15}N-HSQC spectra collected from the mutant Yih1 proteins showing peaks for G96A (blue), G96S (red) (A, C, D) and of wt Yih1 (B). Signals for Gly96 and Gly190 disappear in the spectra collected from the G96A and G96S proteins. Trp135 and Gly134 show two conformations in the spectrum collected from the G96S mutant.

in the solvent increased (1.2, 2.0, 3.0, 4.0, and 5.5 mM) are shown in yellow, green, cyan, sky blue, and blue, respectively. Residues whose relaxation rates were only weakly dependent on the concentration of Gd(DTPA-BMA (smaller than 0.2 s^{-1}·mM^{-1}) are shown in magenta. As the interdomain interface of the second cluster contained residues with diminishing signals (green and blue) in the presence of Gd(DTPA-BMA), indicative of solvent exposure (Fig. S12B), it was rejected. Fig. S12C shows the relative orientation of two ensembles.

In addition, solvent PREs were calculated as the slope of a plot of longitudinal relaxation rate versus concentration of Gd(DTPA-BMA). PREs can reliably be calculated for 101 out of 258 residues making up the protein. A plot (Fig. 6A) of measured PREs (cyan) versus residue number reveals that the peptide N-H moieties in the putative interface have generally lower than average PRE (Fig. 6A). Most of the PREs that cannot be identified are located on flexible loops or on the linker between the two domains, and the corresponding NMR signals are mostly overlapping, missing or quickly disappearing. These exposed residues should have very high PREs, resulting in a higher average PRE which will accentuate the lower PREs of the residues in the putative interface.

We compared solvent PREs that were back-calculated for the full-length structural model (cluster 1) and for the two single domain conformations using a grid-based algorithm [31]. Comparison of these conformations highlight regions with different surface accessibility, most notably in the region around Val98 and Lys173 (Fig. S17A). A plot (Fig. 6A) of calculated and measured (normalized, as described below) solvent PRE versus residue number, shows that the PREs calculated for full-length models (black) parallel the measured PREs

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

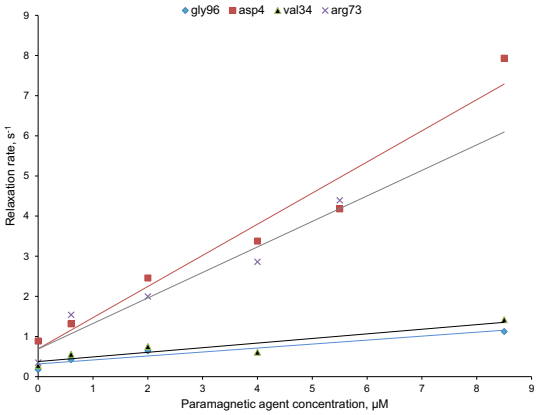


Fig. 5. The dependence of the relaxation rate for Gly96 (blue) on the concentration of the paramagnetic agent Gd(DTPA-BMA). Asp4 (red), Arg73 (gray), Val34 (black). Asp4 and Arg73 are used for comparison, to illustrate the large effect of a paramagnetic agent on the relaxation rate of solvent-exposed residues. Val34 is involved in forming hydrogen bonds between the first and the second strands of the β-sheet and consequently the paramagnetic agent has only a small effect on the relaxation rate. The effect of the paramagnetic agent on the signal from Gly96 signal is similar to that from Val34.

(cyan) much more closely than PREs calculated when the two domains are treated as independent nonassociated entities (orange) especially in the region of the biggest differences between those calculations (Fig. 6B). Experimentally measured solvent PREs can sometimes be higher than calculated, especially in loop regions due to multiple conformations and solvent exchange processes [31]. Four residues in the interdomain interface identified by PISA (two for RWD and two for Ancient domain) illustrate this by showing similar values for measured PREs and PREs calculated for full-length models (Fig. S16). In accordance, the regions of biggest PRE differences calculated for separated domains versus the full-length model are the regions identified by PISA as the interdomain interface (Fig. 6B).

The measured PREs were normalized on the calculated PRE for Ala192, the maximal measured and calculated values (Fig. 6A). For Fig. 6B, further normalization was performed by dividing the differences mentioned above (of PREs calculated for separated domains and for the full-length model) by the experimentally measured PRE. This procedure not only highlights the interdomain interface, but additionally supports our model of ensemble 1, as high measured PREs (indicating the high solvent accessibility) would result in the suggested interdomain interface not highlighted by this procedure (division by big numbers). Our second cluster, ensemble 2, shows different domain interactions. Calculated solvent PREs for this ensemble show critical differences at the interface, as highlighted in Fig. S17B. The region around Lys173 is particularly exposed to the solvent and yields higher theoretical values in contrast to the experimental data and ensemble 1. In contrast, Val98 is rather protected in ensemble 2 and shows lower solvent PREs in disagreement with the experimental data. The position of those residues within the two structural models is shown in Fig. S12C. All PREs and associated residues are tabulated in Table S8.

The resulting ensemble, which is in accordance with all experimental data, is shown in Fig. 7. The interaction between the two domains seems to be at least partially electrostatic, as at neutral pH negatively charged residues in the RWD domain (Glu87, Asp90, Asp102 and Glu106) are close to positively charged residues in the Ancient Domain (Lys173, Arg175 and Lys176), and conversely Arg95 in the RWD and Glu159 in the Ancient Domain are close to each other (Fig. S14). Hydrophobic residues in proximity across the interface include Phe101 and Phe132, Val92 and Leu167, and Val113 and Trp220. An average Ramachandran plot of the models shows that 202 out of 256 residues (79%) are in favored regions, while 238 out of the 256 are in allowed regions (Fig. S15). To improve the quality of the models all three models were further refined in the presence of solvent (water) molecules by XPLOR-NIH using NOEs, RDCs and dihedral angle constraints. The statistics for the final NMR models are given in Table S7.

With respect to the NOE observed between the H(N) of Leu100 and the H(N) of Trp135, the distance calculated (~ 8.6–8.7 Å) lies well outside the standard distance of 5–6 Å for NOEs. A possible explanation is that these two protons are connected through spin diffusion because it has been found that the distance between two protons connected this way can be significantly longer than the 5–6 Å characteristic of a true NOE. The three models (Fig. 7) predict the presence of 87 common NOEs between the RWD and the Ancient domain and although 78 of these were observed, all were ambiguous as the protein has 258 residues and many residues have very similar chemical shifts. Furthermore, all the NOEs were weak, even those predicted to be strong or medium. A possible explanation for this result is some movement in the association of the two domains. Therefore, the problem with identifying interdomain NOEs is not really that NOEs are missing but that they are ambiguous.

Validation of the structural model

The proposed interface comprises a small cluster of hydrophobic residues along with several interdomain

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

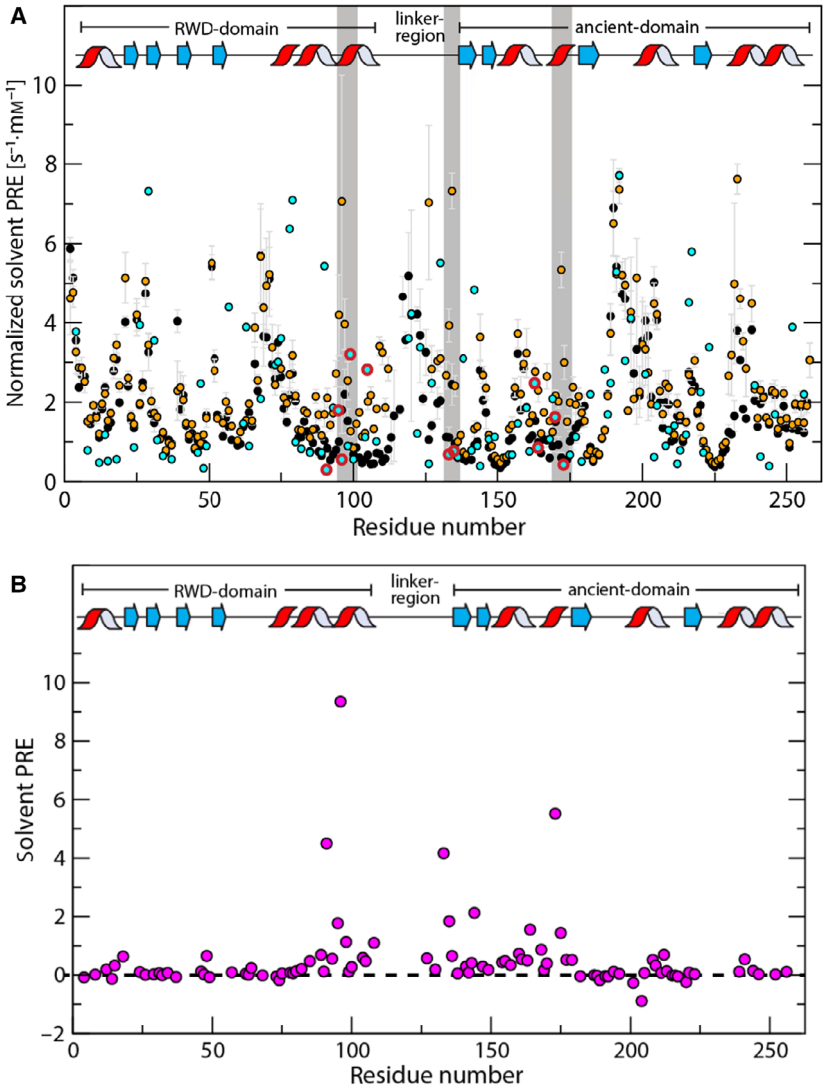


Fig. 6. (A) Plot of experimental (cyan) solvent PREs as a function of residue number for Yih1. Residues identified by PISA26 forming the putative interdomain interface are highlighted in red: 91, 95,96, 99, 105, 133, 135,163, 164, 170, 173. Residues for which reliable PREs could not be obtained are missing. The top panel indicates β-strands and α-helices, respectively. In black, calculated PRE values from the full-length ensemble (ensemble 1) are shown. In orange, the calculated PREs from the free N-terminal and C-terminal domain are shown, values have been normalized to the maximum value of the experimental data. Error bars display the standard deviation from the ensembles. Three regions of high similarity between experimental and calculated values of the full-length model in the domain interface region are highlighted by a dark gray bar. (B) Plot of the average value of PRE calculated for ensemble 1 (black in (A)) subtracted from the PREs calculated for individual domains (orange in (A)) and divided by measured PREs (cyan in A). The three regions that correspond to the interdomain interface are showing high ratios. This procedure would not highlight those regions if the measured PRE would be high, indicating the overall low values of measured PREs in the interdomain interface.

salt bridges. According to PISA [41], the buried area comprises approximately 7% of total surface area, but because of its substantial hydrophilic character has a low free-energy of association as calculated by PISA. However, this ignores the entropic advantage accorded to the association of two tethered domains compared to the entropic penalty resulting from the association of two freely moving independent domains. PPCheck [42] gives the normalized interaction energy per interface residue for the first member of the ensemble as -2.61 kJ·mol^{-1} which fits well within the expected values of -2 to -6 kJ·mol^{-1} obtained from the PDB database. The other models of the ensemble have a low energy (1.7–1.9 kJ·mol^{-1}), indicating that the interface is flexible rather than rigid.

We have eschewed making additional mutants to challenge this interface, due to possible unintended consequences from making such changes. Instead, we have focused our attention on mass spectrometry data of cross-linked protein.

Chemical cross-linking and mass spectrometry

Chemical cross-linking has provided additional validation of our model. Cross-linking of basic residues using standard procedures was attempted, but only intermolecular cross-links were formed. The low concentration of monomers contained no intramolecular cross-links (results not shown). Therefore, we turned our attention to cross-linking acidic residues. The model derived above features a domain interface with many acidic residues (Asp and Glu). In particular, Asp102 and Asp171 are in moderately close proximity to each other (Fig. S18). Carboxylic acid residues are

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

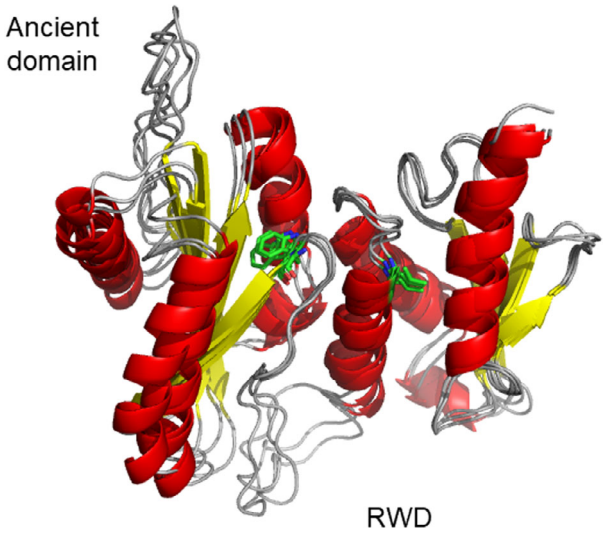


Fig. 7. Ensemble of three models, representing the Yih1 structural model. The structures were calculated with XPLOR-NIH using the SAXS-derived scattering profile, NOEs and RDC data and refined in water with all but SAXS restraints. The selected models are in accordance with all experimental information, including PREs data. Helices are shown in red and beta-strands in yellow. NOE-linked residues 100 and 135 are shown in sticks and colored by element (C green, N blue).

more difficult to cross-link than amino groups, and the only two reagents commercially available for this purpose are ADH (linker length 11.1 Å) and pimelic acid dihydrazide (PDH, linker length 12.3 Å), both of which require the carboxylic acid side chains to be activated by DMTMM. ADH was used as the cross-linking agent, as trials using PDH resulted in unproductive reactions. RP-HPLC of Yih1 treated with ADH at a molar ratio of 1 : 3 showed two distinct peaks (1) and (2) separated from the WT protein (Fig. S19A). SDS/PAGE showed there was minimal intermolecular cross-linking (Fig. S19B), and mass spectrometry showed that peak 1 contained the dimethylated protein with one cross-link (an additional mass of 138.1 amu [34], Fig. S19C) and peak 2 contained dimethylated Yih1 with two and three cross-links, as well as a small amount of nonmethylated Yih1 with three cross-links (Fig. S19D).

In an attempt to identify which amino acids were involved in the formation of the cross-links, cross-linked Yih1 (containing up to three cross-links), plus non-cross-linked control samples, were first subjected to trypsin digestion and subsequently to a double digestion using trypsin followed by chymotrypsin. The peptide coverage for the non-cross-linked protein was 81%, which reduced to 75% for the sample with one or two cross-links (Fig. S18). Although the unmatched peptides did not significantly change between the two samples, there were minor changes around Asp102 (in the RWD domain) and Glu249 (in the Ancient Domain), which may be linked to produce a larger peptide, or one of them may be linked to a different partner. Peptide 166–173 or sometimes 170–173 was consistently not seen in either sample as was peptide 43–73 (Fig. S20). The former peptide should be cleaved into small peptides that may not bear a positive charge because of their sequence, while the latter peptide would not be cleaved by chymotrypsin due to the presence of proline residues and due to its amino acid composition, may not form a positively charged ion. Although the distance between the side chains of Asp102 and Asp171 (in the range 6.3–9.9 Å) is smaller than the length of the cross-linker itself, a link between them is not implausible as the two domains are known from SAXS and NMR data to be loosely coupled. In order to confirm the existence of at least one cross-link, the digestion products of multiple cross-linking reactions were combined and subjected to SEC. Mass spectrometry analysis of 140 individual, 0.5 mL fractions collected from the Biogel P2 column showed that only one fraction, 43, contained a peak with a mass that matched the expected mass for a cross-link between residues 102–122 and residues 166–173. The peak at m/z 997.4863 represents the cross-linked peptide ^{101}FDF-^{170}TDSK with the neutral loss of water (Fig. S21) and fragmentation produced the ions expected (Table S9). No fractions were found that contained masses consistent with a cross-link between D102 and E249 or D257 or between E56, E57, or E63 in peptide 43–73 and D171 (Fig. S20). Finally, although the far-UV region of the CD spectrum of cross-linked protein was unchanged (Fig. S22), the near-UV region of the cross-linked protein changed, especially the regions around 260 and 280 nm consistent with the expected perturbations of the environments of Phe101 and Phe103 (Fig. S23), which sit adjacent to the cross-linked Asp102. Furthermore, in the structure modeled for the native Yih1, Asp 102 and Asp 171 are near Trp135, which lies close to the interface. That the overall shape of the near-UV spectrum remains similar for wild-type and cross-linked Yih1 suggests that not only the secondary structure but also the tertiary structure of the cross-linked protein is maintained.

Discussion

Similar structures in PDB database and implications for IMPACT proteins from different species

A DALI [43] search did not give any hits for the combination of the two domains; thus, the combination

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

observed here is unique to Yih1/IMPACT proteins. As expected, the closest homologues to the Ancient Domain of Yih1 were the Ancient Domains of *T. thermophilus* hypothetical protein (PDB-ID: 2cve) with RMSD of 2.2 Å and 34.8% identity for structural superposition of 112 out of 122 (Yih1) residues, and of YigZ protein from *E. coli* (PDB-ID: 1vi7) with RMSD of 2.6 Å and 23.1% identity for superposition of 91/122 (Yih1) residues. The best matches for the RWD domain of Yih1 were to the RWD domains of the human protein-1 (PDB-ID: 2ebm) with RMSD 2.8 Å and 21.9% identity for 96/112 (Yih1) residues and of mouse Gcn2 (PDB-ID: 1ukx) with RMSD 2.4 Å and 18.6% identity for 97/112 (Yih1) residues. Remarkably, the *S. cerevisiae* RWD domains of Yih1 and Gcn2 (PDB-ID: 2yz0) show low sequence identity of just 14.0% for rather close structural similarity [RMSD 2.8 Å for superposition of 100 out of 112 (Yih1) residues]. This reflects either an early evolutionary separation of these RWD domains or strong selection pressure in *S. cerevisiae* for differentiation to ensure fidelity of function. The RWD domain of human protein-3 (PDB-ID: 2ebk), related to human protein-1 also shows remarkable structural similarity to the RWD domain of Yih1 (RMSD 2.7 Å for superposition of 95 out of 112 residues) despite sequence identity of just 9.5%.

Glu87 and Glu106 in the RWD domain and Arg175 and Lys173 in the Ancient Domain of Yih1 are able to form interdomain salt bridges in our model and could therefore be involved in controlling the relative orientation of the two domains and, as a consequence, Yih1 function. If this is the case, these salt bridges should be conserved among Yih1/IMPACT orthologues. While Lys173 and Glu87 are conserved in IMPACT proteins from different species [7], Glu106 and Arg175 are not. Appropriately charged residues are, however, found at locations similar to those of Glu106 and Arg173, that is, on helix 2 of the RWD domain and close to the end of helix 1 of the Ancient Domain [7], potentially maintaining the electrostatic interactions between the domains of IMPACT proteins from different species. Homology modeling of human IMPACT based on our Yih1 model (Fig. S24) and calculated with YASARA [35,44] supports this hypothesis. Human IMPACT and Yih1 share 36.2% sequence identity over 207 residues.

Mechanism of Yih1 interactions

In our model, Asp102 and/or Glu106, which have been previously identified as being important for the interaction of Gcn1 with Yih1 [7], are buried in the domain interface of Yih1, suggesting that for Gcn1-Yih1 interaction to occur there has to be some domain movement in Yih1 (Fig. S14). This would not be unusual, as in several studied interactomes, large-amplitude motions between protein domains play important roles in biomolecular function [45–48]. The E87A:D90A double substitution in Yih1 leads to a distinct phenotype in yeast cells, proposed to be caused by a stronger Yih1-Gcn1 interaction, resulting in an increased inhibition of Gcn2 [7]. In the absence of a structural model for the entire Yih1 protein, however, the reason for this phenotype was not clear at that time. Because Glu87 is positioned to make an interdomain salt bridge with Arg175, it could be argued that mutation to Ala should weaken the interdomain interaction, allowing domain movements to expose Asp102 and Glu106, thus allowing stronger binding of Gcn1 to Yih1. It could also be argued that such a mutation might enhance the hydrophobic interaction of Ala with the $(CH_2)_3$ portion of the Arg side chain. In any event, destabilization of the interaction between the two domains exposes Asp102 or Glu106 of Yih1 to Arg2259 of Gcn1, potentially allowing the formation of a salt bridge and thus stabilizing the interaction of Yih1 with Gcn1.

The E87A:D90A double-mutant also results in Yih1 making stronger interactions with actin and Cdc28 [4,5,7,9]. Furthermore, removing residues 1–67 (leaving the remnants of the RWD domain mostly likely unfolded) results in a very strong interaction between actin and Yih1 [4], indicating that actin may interact exclusively with the Ancient Domain of Yih1. The extension of the Ancient Domain (residues 127–136) and part of the linker (residues 113–126) may be still required for interaction with other physiological partners, since a Yih1 fragment containing the originally proposed Yih1 Ancient Domain (amino acids 133–258) does not co-precipitate actin or Cdc28 *in vivo* [7,9].

Taken together, these results suggest that Yih1 in its closed conformation may not bind to any of its verified partners but a partner with a higher affinity for one of the domains can select the initially low population of the open conformation and stabilize it by binding to one of the domains. SAXS and solvent PRE data strongly indicate that the average population is indeed closed, but evidence of some flexibility (RDC, weak NOE signals, and SAXS) suggests the possibility of an equilibrium between the open and closed conformations. To determine the exact populations of these conformations would require considerable work. The relative strengths of the interdomain interaction versus the interaction of the open conformation of Yih1 with actin or Gcn1 appear to be functionally important. How then might the strength of the interdomain interaction, and the strength of interaction of Yih1 with its potential partners, be modulated?

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

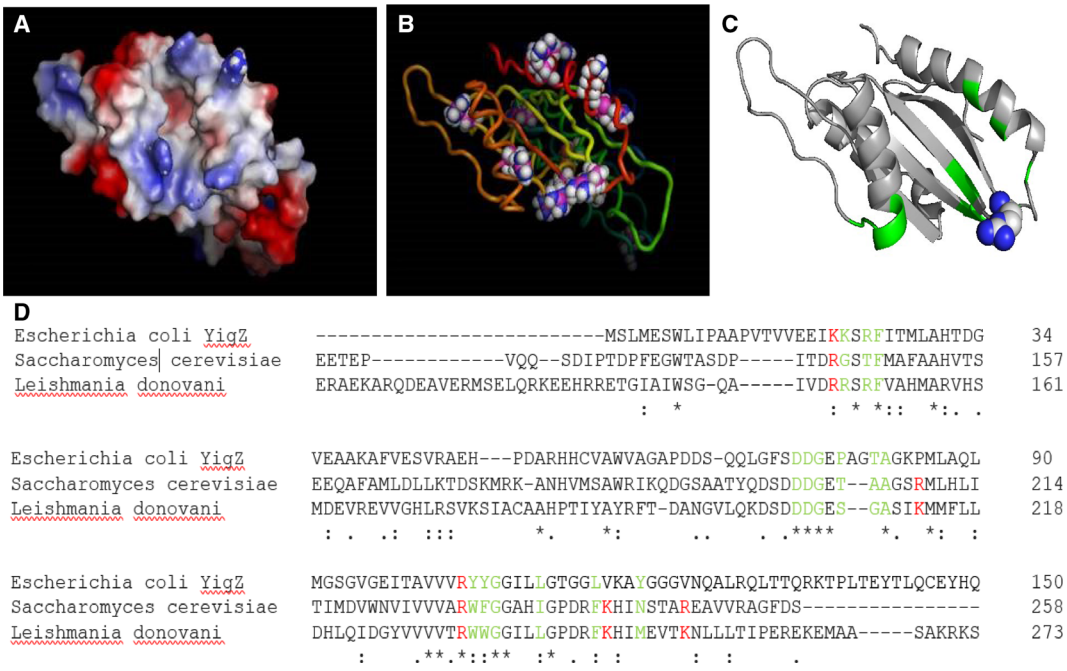


Fig. 8. The predicted RNA binding site of Yih1: (A) Electrostatic surface. (B) Ribbon diagram showing the arginines and lysines around the deep groove. (C) Cartoon diagram of the Ancient Domain with region predicted to bind RNA in YIGZ protein shown in green [36]. The dimethylated Arg144 residue is shown as spheres. (D) Sequence alignment (Clustal Omega) around predicted RNA binding site (green) of YigZ protein and IMPACT proteins from *S. cerevisiae* (Yih1) and from *Leishmania donovani*. Arg and Lys residues surrounding the putative RNA binding groove are highlighted in red.

Possible functional roles of methylation

An overarching regulatory mechanism, such as post-translational modification, may shift the balance point between the closed and open states of Yih1, and thereby modulate its binding with other partner molecules. The physiological roles of protein methylation have been widely studied. Methylation is involved in signal transduction, mRNA splicing, transcriptional control, DNA repair, protein translocation, and mitochondrion energy metabolism [49,50]. Although the Yih1 used in this study is a product of heterologous expression, the observed dimethylation of Arg144 or Lys36 points to a possible mechanism that could be used by *S. cerevisiae* for shifting the balance between the open and closed conformations thus influencing key cellular functions.

Previously we had proposed that the Ancient Domain has a binding interface for yet unknown binding partner (s) [7]. In the Ancient Domain, there is an absolutely conserved, highly negatively charged region ^{200}DDGE and a strongly hydrophobic region containing one positive Arg, (^{224}V/A/I)(V/M)V(A/T/S)**R**(W/Y)(F/Y)GG (A/I/V). This part of Ancient Domain is not affected, at least directly, by the interdomain interactions within Yih1. We can speculate, that if these motifs are involved in binding yet-to-be-discovered binding partners, their effect on interdomain interactions within Yih1, is likely to be *via* entropically based mechanisms, as is increasingly found in allosteric regulation [51–53]. The most likely binding partner for this region is DNA or RNA, consistent with the recently discovered nuclease activity of IMPACT proteins [54,55]. The nuclease activity of the fungus *Chaetomium thermophilum* IMPACT homolog [54] and the predicted nucleic acid binding interface of YigZ protein [36] can be mapped to this region (Fig. 8). The corresponding surface for Yih1 (Fig. 8) shows a deep groove surrounded by positive side chains, including Arg144, that would be able to interact with negatively charged phosphate backbones of oligonucleotides. Arg144 was identified by mass spectrometry as being partially dimethylated. Dimethylation of this arginine could then regulate DNA/RNA binding and/or nuclease activity of Yih1, and potentially also its interactions with the ribosome.

Conclusions and Perspectives

We have determined the NMR structure of each domain of Yih1, previously identified as an N-terminal RWD domain and a C-terminal Ancient Domain. Combining different biophysical methods has then

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

enabled us to generate a structural model of full-length Yih1, which adopts a closed but somewhat flexible conformation. These results provide greater insight into how Yih1—through interactions with Gcn1—controls the activity of Gcn2 in modulating the rate of protein synthesis. In particular, residues identified as important for interaction with Gcn1 are largely buried or sequestered into salt bridges and in interacting with Gcn1, the Yih1 must adopt an open conformation.

Significantly, features of the structural model of Yih1, such as the groove lined with positively charged residues, one of which (Arg144) was found to be dimethylated, point to a raft of studies focused on elaborating the cascade of interactions in which Yih1 appears to play a central role. These include key biological processes such as the starvation response, development, proper brain function and immunity. The immediate challenge is to identify potential methylation of Yih1 in its native environment and, if observed, to determine (a) the specific physiological stimulus (or stimuli) that cause methylation and (b) whether it is the methylated or non-methylated form of the protein that controls affinity of Yih1 for Gcn1 and other binding partners, and how this affects the many functions of Gcn2.

Acknowledgements

We would like to thank Massey NMR and mass spectrometry facilities and Australian Synchrotron and SAXS beam-line staff. Accession numbers: PDB-ID: 6u1l, 6u1o; bmrb: 30375. We also thank the editor and referees for constructive criticism.

Author contributions

EH, GBJ, ES and GEN designed the research and wrote the manuscript. NB and TSL established the protein purification protocol, including isotopically labelled protein. TSL purified wildtype and mutant proteins. AKG and PJBE collected and processed NMR spectra for assignment and structure calculation. EH performed the NMR assignment and structure calculation. GBJ collected and analysed SAXS data. BM collected and with EH analysed RDC data. SH collected PRE data; EH and SH did preliminary interpretation; CG analysed PRE data. Y-HT conducted crosslinking studies. Y-HT, TSL and GEN collected and analysed mass spectrometry data.

References

1 Castilho BA, Shanmugam R, Silva RC, Ramesh R, Himme BM and Sattlegger E (1843) Keeping the eIF2 alpha kinase Gcn2 in check. *Biochim Biophys Acta* **1843**, 1948–1968.

2 Silva RC, Sattlegger E and Castilho BA (2016) Perturbations in actin dynamics reconfigure protein complexes that modulate GCN2 activity and promote an eIF2 response. *J Cell Sci* **129**, 4521–4533.

3 Roffe M, Hajj GN, Azevedo HF, Alves VS and Castilho BA (2013) IMPACT is a developmentally regulated protein in neurons that opposes the eukaryotic initiation factor 2alpha kinase GCN2 in the modulation of neurite outgrowth. *J Biol Chem* **288**, 10860–10869.

4 Sattlegger E, Swanson MJ, Ashcraft EA, Jennings JL, Fekete RA, Link AJ and Hinnebusch AG (2004) YIH1 is an actin-binding protein that inhibits protein kinase GCN2 and impairs general amino acid control when overexpressed. *J Biol Chem* **279**, 29952–29962.

5 Sattlegger E and Hinnebusch AG (2000) Separate domains in GCN1 for binding protein kinase GCN2 and ribosomes are required for GCN2 activation in amino acid-starved cells. *EMBO J* **19**, 6622–6633.

6 Marton MJ, Vazquez de Aldana CR, Qiu H, Chakraburtty K and Hinnebusch AG (1997) Evidence that GCN1 and GCN20, translational regulators of GCN4, function on elongating ribosomes in activation of eIF2alpha kinase GCN2. *Mol Cell Biol* **17**, 4474–4489.

7 Sattlegger E, Barbosa JA, Moraes MC, Martins RM, Hinnebusch AG and Castilho BA (2011) Gcn1 and actin binding to Yih1: implications for activation of the eIF2 kinase GCN2. *J Biol Chem* **286**, 10341–10355.

8 Sattlegger E, Chernova TA, Gogoi NM, Pillai IV, Chernoff YO and Munn AL (2014) Yeast studies reveal moonlighting functions of the ancient actin cytoskeleton. *IUBMB Life* **66**, 538–545.

9 Silva RC, Dautel M, Di Genova BM, Amberg DC, Castilho BA and Sattlegger E (2015) The Gcn2 regulator Yih1 interacts with the cyclin dependent kinase Cdc28 and promotes cell cycle progression through G2/M in budding yeast. *PLoS One* **10**, e0131070.

10 Waller T, Lee SJ and Sattlegger E (2012) Evidence that Yih1 resides in a complex with ribosomes. *FEBS J* **279**, 1761–1776.

11 Doerks T, Copley RR, Schultz J, Ponting CP and Bork P (2002) Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res* **12**, 47–56.

12 Kubota H, Sakaki Y and Ito T (2000) GI domain-mediated association of the eukaryotic initiation factor 2alpha kinase GCN2 with its activator GCN1 is required for general amino acid control in budding yeast. *J Biol Chem* **275**, 20243–20246.

13 Nameki N, Yoneyama M, Koshiba S, Tochio N, Inoue M, Seki E, Matsuda T, Tomo Y, Harada T, Saito K

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

et al. (2004) Solution structure of the RWD domain of the mouse GCN2 protein. *Protein Sci* **13**, 2089–2100.
14 Lang EJ, Cross PJ, Mittelstadt G, Jameson GB and Parker EJ (2014) Allosteric ACTion: the varied ACT domains regulating enzymes of amino-acid metabolism. *Curr Opin Struct Biol* **29**, 102–111.
15 Hagiwara Y, Hirai M, Nishiyama K, Kanazawa I, Ueda T, Sakaki Y and Ito T (1997) Screening for imprinted genes by allelic message display: identification of a paternally expressed gene impact on mouse chromosome 18. *Proc Natl Acad Sci USA* **94**, 9249–9254.
16 Chen KM, Harjes E, Gross PJ, Fahmy A, Lu Y, Shindo K, Harris RS and Matsuo H (2008) Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116–119.
17 Zhang O, Kay LE, Olivier JP and Forman-Kay J (1994) Backbone 1H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J Biomol NMR* **4**, 845–858.
18 Ottiger M, Delaglio F and Bax A (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* **131**, 373–378.
19 Zweckstetter M (2008) NMR: prediction of molecular alignment from structure using the PALES software. *Nat Protocols* **3**, 679–690.
20 Harjes E, Kitamura A, Zhao W, Morais MC, Jardine PJ, Grimes S and Matsuo H (2012) Structure of the RNA claw of the DNA packaging motor of bacteriophage Phi29. *Nucleic Acids Res* **40**, 9953–9963.
21 Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS *et al*. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905–921.
22 Herrmann T, Güntert P and Wüthrich K (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* **24**, 171–189.
23 Herrmann T, Guntert P and Wuthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* **319**, 209–227.
24 Cross PJ, Dobson RC, Patchett ML and Parker EJ (2011) Tyrosine latching of a regulatory gate affords allosteric control of aromatic amino acid biosynthesis. *J Biol Chem* **286**, 10216–10224.
25 Schwieters CD, Kuszewski JJ, Tjandra N and Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* **160**, 65–73.
26 Schwieters CD and Clore GM (2014) Using small angle solution scattering data in Xplor-NIH structure calculations. *Prog Nucl Magn Reson Spectrosc* **80**, 1–11.
27 Hocking HG, Zangger K and Madl T (2013) Studying the structure and dynamics of biomolecules by using soluble paramagnetic probes. *ChemPhysChem* **14**, 3082–3094.
28 Respondek M, Madl T, Gobl C, Golser R and Zangger K (2007) Mapping the orientation of helices in micelle-bound peptides by paramagnetic relaxation waves. *J Am Chem Soc* **129**, 5228–5234.
29 Keller R (2004) Optimizing the Process of Nuclear Magnetic Resonance Spectrum Analysis and Computer Aided Resonance Assignment. ETH, Zürich.
30 Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J and Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696.
31 Hartlmüller C, Göbl C and Madl T (2016) Prediction of protein structure using surface accessibility data. *Angew Chem Int Ed* **55**, 11970–11974.
32 Krissinel E and Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D* **60**, 2256–2268.
33 Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501.
34 Leitner A, Joachimiak LA, Unverdorben P, Walzthoeni T, Frydman J, Förster F and Aebersold R (2014) Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc Natl Acad Sci USA* **111**, 9455–9460.
35 Krieger E, Koraimann G and Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA - A self-parameterizing force field. *Proteins* **47**, 393–402.
36 Park F, Gajiwala K, Eroshkina G, Furlong E, He D, Batiyenko Y, Romero R, Christopher J, Badger J, Hendle J *et al*. (2004) Crystal structure of YIGZ, a conserved hypothetical protein from *Escherichia coli* k12 with a novel fold. *Proteins* **55**, 775–777.
37 Shen Y, Delaglio F, Cornilescu G and Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213–223.
38 Konarev PV, Petoukhov MV and Svergun DI (2001) MASSHA - a graphics system for rigid-body modelling of macromolecular complexes against solution scattering data. *J Appl Crystallogr* **34**, 527–532.
39 Pintacuda G and Otting G (2002) Identification of protein surfaces by NMR measurements with a paramagnetic Gd(III) chelate. *J Am Chem Soc* **124**, 372–373.
40 Walsh JD, Meier K, Ishima R and Gronenborn AM (2010) NMR studies on domain diffusion and alignment in modular GB1 repeats. *Biophys J* **99**, 2636–2646.

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

41 Krissinel E and Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774–797.
42 Sukhwal A and Sowdhamini R (2013) Oligomerisation status and evolutionary conservation of interfaces of protein structural domain superfamilies. *Mol Biosyst* **9**, 1652–1661.
43 Holm L, Kääriäinen S, Rosenström P and Schenkel A (2008) Searching protein structure databases with DaliLite vol 3. *Bioinformatics* **24**, 2780–2781.
44 Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D and Karplus K (2009) Improving physical realism, stereochemistry and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins* **77**, 114–122.
45 Bahar I, Chennubhotla C and Tobi D (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr Opin Struct Biol* **17**, 633–640.
46 Smock RG and Gierasch LM (2009) Sending signals dynamically. *Science* **324**, 198–203.
47 Tzeng SR and Kalodimos CG (2011) Protein dynamics and allostery: an NMR view. *Curr Opin Struct Biol* **21**, 62–67.
48 Mackereth CD and Sattler M (2012) Dynamics in multi-domain protein recognition of RNA. *Curr Opin Struct Biol* **22**, 287–296.
49 Sha L, Daitoku H, Araoi S, Kaneko Y, Takahashi Y, Kako K and Fukamizu A (2017) Asymmetric arginine dimethylation modulates mitochondrial energy metabolism and homeostasis in *Caenorhabditis elegans*. *Mol Cell Biol* **37**, e00504-16.
50 Bedford MT and Clarke SG (2009) Protein arginine methylation in mammals: who, what, and why. *Mol Cell* **33**, 1–13.
51 Lang EJ, Heyes LC, Jameson GB and Parker EJ (2016) Calculated pKa variations expose dynamic allosteric communication networks. *J Am Chem Soc* **138**, 2036–2045.
52 Kornev AP (2018) Self-organization, entropy and allostery. *Biochem Soc Trans* **46**, 587–597.
53 Saavedra HG, Wrabl JO, Anderson JA, Li J and Hilser VJ (2018) Dynamic allostery can drive cold adaptation in enzymes. *Nature* **558**, 324–328.
54 Zamora Caballero S (2016) *Structural and Functional Characterization of IMPACT Proteins: A Novel Nuclease Family*. CSIC - Instituto de Biomedicina de Valencia (IBV) Universidad de Valencia.
55 Rao SJ, Shukla E, Bhatia V, Lohiya B, Gaikwad SM, Kar A and Pal JK (2018) The Leishmania donovani IMPACT-like protein possesses non-specific nuclease activity. *Int J Biol Macromol* **119**, 962–973.
56 Schneidman-Duhovny D, Hammel M and Sali A (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res* **38**, W540–544.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Instrument configuration for mass spectrometry analysis of cross-linked Yih1.

Table S2. Instrument configuration for mass spectrometry analysis of Yih1 cross-links

Table S3. Method parameter settings for mass spectrometry analysis of Yih1 cross-links.

Table S4. Search parameters for MSMS.

Table S5. Compactness of backbone (BB) and side chain assignment (SC) for Yih1 protein.

Table S6. Table of chemical shift perturbations (CSPs) for wild-type Yih1 versus G96S and G96A mutants and versus I127A and I235V mutants.

Table S7. (A) Summary of NMR structural parameters for the ensemble of three structures representing Yih1 full-length structural model (pdb code. 6BQI) after explicit water refinement. (B): Summary of NMR structural parameters for RWD (pdb code: 6U1L) and Ancient Domain (pdb code: 6U1O) as separate structures calculated with the connection (linker is present in the calculation, no SAXS, no interdomain NOEs).

Table S8. Paramagnetic relaxation enhancement (PRE) data calculated and measured.

Table S9. Expected fragment ions for from the singly charged cross-linked precursor 997.4836 amu.

Fig. S1. Chromatogram of Yih1 anion exchange chromatography.

Fig. S2. Size exclusion chromatogram (SEC) of Yih1 from anion exchange chromatography.

Fig. S3. SDS/PAGE of SEC fractions for SEC of Yih1.

Fig. S4. Deconvoluted mass spectrum of WT Yih1 after SEC, showing the presence of native and modified Yih1 in the sample.

Fig. S5. MSMS spectrum of the peak $(M + 2H)^{2+}$= 622.367297 amu.

Fig. S6. MSMS spectrum of the peak $(M + 2H)^{2+}$= 502.257391 amu.

Fig. S7. NOEs between HD2 of Ile127 and HN of Arg239 as well as HN of Phe240 from helix 3 of the ancient domain: strips from 3D-15N-NOESY-HSQC spectra.

Fig. S8. Cartoon diagram of the representative model of the final ensemble illustrating the effect of mutations in one domain on the chemical shifts of the other domain.

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Fig. S9. Plot of chemical shift perturbations (CSPs) for wild-type Yih1 versus (A) G96S (orange) and G96A (blue) mutants; and (B) I127A (blue) and I235V (orange) mutants.

Fig. S10. Gly96 (magenta) is found on a surface loop of the RWD domain.

Fig. S11. Inter-domain NOE between HN of Leu100 and HE1 of Trp135 and HN of Trp135 seen in three NOESY spectra.

Fig. S12. Solvent accessibility data for ensemble 1 (A), and ensemble 2 (B). Amino acids with resonances disappearing in 1.2 mM Gd(DTPA-BMA) are highlighted in yellow; those disappearing in 2 mM Gd(DTPA-BMA) are green; those disappearing in 3 mM Gd (DTPA-BMA) are cyan, those disappearing in 4 mM Gd(DTPA-BMA) are sky blue and those disappearing in 5.5 mM Gd(DTPA-BMA) are blue. Residues showing weak dependence of relaxation rate on the concentration of the paramagnetic Gd(DTPA-BMA) (smaller than 0.2 s-1· mM–1) are highlighted in magenta. Gly96 is shown as spheres. (C) Models of the ensemble 1 (cyan, semi-transparent) and the ensemble 2 (orange), both structure are overlaid on RWD and the difference in the position of Ancient domain is clearly to see. The amide proton of residue Lys173 is depicted as a sphere and shows reduced accessibility (part of inter-domain interface) in the ensemble 1 and higher solvent accessibility in the ensemble 2. The other highlighted residue, Val98, shows higher accessibility in the ensemble 1 conformation and lower calculated values in the ensemble 2.

Fig. S13. (A) Plots of observed versus calculated RDCs for the two ensembles. Ensemble 1 (upper panel, RWD blue and Ancient Domain yellow) was selected for further development as the structural model for full-length Yih1. (B) R1 relaxation rate plotted against amino acid number.(C) SAXS fits of representative model of ensemble 1 (left) and ensemble 2 (right)

Fig. S14. A representative structure of ensemble 1 showing Glu87, Asp90, Asp102, Glu106 and Glu159 as red, and Arg95, Lys173, Arg175 and Lys176 as blue sticks.

Fig. S15. Ramachandran Plot (calculated with MolProbity, http://molprobity.biochem.duke.edu) of three Yih1 models of ensemble 1.

Fig. S16. Full-length Yih1 structural models highlighting the examples of residues in the inter-domain interface, identified by PISA.

Fig. S17. (A) Comparison of calculated solvent PREs from lowest energy structures of the individual domains (RWD-domain in cyan and ancient-domain in magenta) and the full-length ensemble 1 structural model (black), grey error bars indicate the standard deviation derived from the ensembles. Dark grey bars highlight regions of the interface showing the highest differences between the ensembles. (B) Comparison of calculated solvent PREs from ensemble1 (black) and ensemble2 (grey), to the measured solvent PREs (cyan); standard deviation is indicated by grey error bars. The linker region is highlighted by a grey bar. The two dark grey bar highlight areas where the ensemble1 correlates better than the ensemble2 conformation.

Fig. S18. Representative structural model (ensemble 1) with the Asp102 and Asp171 shown as red sticks.

Fig. S19. (A) RP-HPLC chromatogram showing optimized ADH cross-linking conditions (red line, Yih1 to ADH molar ratio of 1 : 3 in HEPES buffer, pH 7.0 at room temperature and stopped after 1 h); blue line, WT; black line, initial ADH cross-linking trial. (B) SDS/PAGE of red peaks 1 and 2 from RP-HPLC. (C) Deconvoluted mass spectrum of red peak 1. (D) Deconvoluted mass spectrum of red peak 2, showing the addition of ADH cross-links to Yih1. +1, +2 and + 3 refer to the addition of ADH cross-links (+138 amu). A is wild-type Yih1 and B is dimethylated Yih1.

Fig. S20. Sequence of Yih1 showing the peptides identified by mass spectrometry (red bold) following a double digestion, first with trypsin followed by chymotrypsin.

Fig. S21. Mass spectrum of peak 43.

Fig. S22. Far-UV CD spectra of wild-type (A) and cross-linked Yih1 (B), showing preservation of secondary structure.

Fig. S23. Near-UV CD spectra of wild-type (A) and cross-linked Yih1 (B), showing differences in tertiary structure.

Fig. S24. Cartoon diagram of homology model of human IMPACT protein, highlighting electrostatic interdomain interaction.

18733468, 2021, 3, Downloaded from https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13990 by CochraneChina, Wiley Online Library on [02/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License