NMR structure of a non-conjugatable, ADP-ribosylation associated, ubiquitin-like domain from *Tetrahymena thermophila* polyubiquitin locus

Valerio Chiarinia,b, Helena Tossavainib, Vivek Sharmaa,c, Gianni Colottid

a Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, Viikinkaari 1, P.O. Box 65, FI-00014 Helsinki, Finland
b Department of Chemistry, Nanoscience Center, University of Jyväskyla, P.O. Box 35, FI-40014 Jyväskyla, Finland
c Department of Physics, University of Helsinki, P.O. Box 64, FI-00014 Helsinki, Finland
d Institute of Molecular Biology and Pathology, Italian National Research Council (IBPM CNR), Department of Biochemical Sciences, Sapienza University, P.le A. Moro 5, 00185 Rome, Italy

**Abstract**

Background: Ubiquitin-like domains (UbLs), in addition to being post-translationally conjugated to the target through the E1-E2-E3 enzymatic cascade, can be translated as a part of the protein they ought to regulate. As integral UbLs coexist with the rest of the protein, their structural properties can differ from canonical ubiquitin, depending on the protein context and how they interact with it. In this work, we investigate *T*. *th*-ubl5, a UbL present in a polyubiquitin locus of *Tetrahymena thermophila*, which is integral to an ADP-ribosyl transferase protein. Only one other co-occurrence of these two domains within the same protein has been reported.

Methods: NMR, multiple sequence alignment, MD simulations and SPR have been used to characterize the structure of *T*. *th*-ubl5, identify putative binders and experimentally test the interaction, respectively.

Results: Molecular dynamics simulations showed that *T*. *th*-ubl5 is unable to bind the proteasome like ubiquitin due to the lack of the conserved hydrophobic patch. Of other integral UbLs identified by structural and sequence alignment, *T*. *th*-ubl5 showed high structural and sequence resemblance with the Ras-binding epitope of FERM UbLs. SPR experiments confirmed that a strong and specific interaction occurs between *T*. *th*-ubl5 and *T*. *th*-Ras.

Conclusion: Data indicate that *T*. *th*-ubl5 does not interact with the proteasome like ubiquitin but acts as a decoy for the recruitment of Ras protein by the ADP-ribosyl transferase domain.

General significance: Mono-ADP-ribosylation of Ras proteins is known as a prerogative of bacterial toxins. *T*. *th*-ubl5 mediated recruitment of Ras highlights the possibility of an unprecedented post-translational modification with interesting implication for signalling pathways.

**1. Introduction**

Ubiquitination is a protein modification process where the C-terminus of ubiquitin is enzymatically attached to a lysine side chain (-NH$_3^+$ group) of the target protein. This process is catalysed by the E1-E2-E3 enzymatic cascade [1]. The main function of this modification is to promote the proteasomal degradation of the target protein. Several other ubiquitin-like domains (UbLs) have been identified and grouped, and together with ubiquitin they form the ubiquitin superfamily. The ubiquitin-like fold, also called the β-grasp fold, is one of the most represented three-dimensional structures found in the protein universe [2]. Despite its high structural conservation across various protein families, the sequence similarity is low [3].

Members of the ubiquitin superfamily can be divided into two groups based on whether they are post-translationally conjugated or not. Ubiquitin-like modifiers (ULMs) comprise different protein domains, which become attached to their target protein by enzymes similar to the enzymatic cascade operating on ubiquitin [4]. This group includes SUMO, NEDD8, ISG15, APG8, FAT10, URM1, Ufm1 and Hub1 [5]. Contrary to ubiquitin, these domains are not involved in protein

**Abbreviations:** H. *sapiens*, Homo *sapiens*; *T*. *th*, *Tetrahymena thermophila*; UbL, ubiquitin-like domain; UDP, UbL-containing proteins; ULM, Ubiquitin-like modifiers; UIM, Ubiquitin interacting motif; BII, Bacterial intein like; ART, ADP-ribosyl transferase; MD, Molecular dynamics; SPR, Surface plasmon resonance; NMR, Nuclear magnetic resonance

⁎ Corresponding author.

⁎⁎ Corresponding author at: Institute of Molecular and Pathology, Italian National Research Council (IBPM CNR), c/o Department of Biochemical Sciences, Sapienza University, P.le A. Moro 5, 00185 Rome, Italy.

E-mail addresses: valerio.chiarini@helsinki.fi (V. Chiarini), gianni.colotti@uniroma1.it (G. Colotti).

https://doi.org/10.1016/j.bbagen.2019.01.014

Received 8 October 2018; Received in revised form 27 December 2018; Accepted 22 January 2019

Available online 25 January 2019

0304-4165/ © 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
degradation but exert other functions, often related to signalling and trafficking pathways. ULMs are translated either as precursors or single domains, and they share a conserved R-G-G sequence in their C-terminus, required for their maturation and conjugation. In addition to being found as independently expressed domains, UBLs can also be part of larger proteins. Proteins containing UBLs are called ubiquitin-like domain proteins (UDPs) [6]. UBLs belonging to this group are neither processed nor conjugated but instead are integral part of the “host protein” since its translation. Because of their sequence similarity with ubiquitin, integral UBLs are often able to interact with a plethora of proteins, including the proteasome, with which UBLs interact through the key residues, known as those forming the “hydrophobic patch”, equivalent to residues L8, 144 and V70 of ubiquitin. This interaction, however, does not result in the degradation of the host protein [7]. In fact, in some cases, like the deubiquitinase USP14, the interaction of the UBL domain with the proteasome is critical for efficient catalysis [8].

Both the E3-ligase parkin and the UV- excision repair protein Rad23 are also recruited onto the proteasome upon interaction between their UBLs and the Rpn10 proteasomal subunit [9] [10] [11] and mutations or deletions of the UBL are associated with the loss of the host protein function causing pathological conditions, such as the autosomal recessive juvenile Parkinsonism [12].

Interestingly, S. cerevisiae Rad23 (UNIPROT_P32628) functional rescue was observed for the mutant lacking the UBL when the missing domain was replaced with ubiquitin [9]. This evidence indicates that the conservation of the hydrophobic patch is sufficient to guarantee the function of the UBL while variations of the conservation of the hydrophobic patch is sufficient to guarantee the rescue was observed for the mutant lacking the UbL when the missing domain was replaced with ubiquitin [9]. This evidence indicates that the conservation of the hydrophobic patch is sufficient to guarantee the function of the UBL while variations of the conservation of the hydrophobic patch is sufficient to guarantee the rescue was observed for the mutant lacking the UbL when the missing domain was replaced with ubiquitin [9]. This evidence indicates that the conservation of the hydrophobic patch is sufficient to guarantee the function of the UBL while variations of the conservation of the hydrophobic patch is sufficient to guarantee the rescue was observed for the mutant lacking the UbL when the missing domain was replaced with ubiquitin [9].

Alternative functions for UBLs have been established as well. Finley et al. [13] described the chaperone-like function observed for UBL of ribosomal proteins, and their importance in promoting the ribosome assembly. In parkin, the UBL maintains the host protein in an ‘idle’ state until both the phosphorylation of UBL and the interaction of parkin with the phosphorylated ubiquitin have occurred [12] [14] [15].

Integral UBLs featuring the conserved hydrophobic surface use it to exert the regulative function by interacting with either the host protein or the foreign protein(s), or both. Thus, integral UBLs lacking the patch are unable to establish similar interactions with the ubiquitin-interacting motif (UIM) [16], ruling out the possibility to be recognized by most ubiquitin-binding proteins, although not necessarily by their host protein.

In this work, we describe the NMR solution structure of the T.th-ubl5 domain of Tetrahymena thermophyla BUBL1 (UBL-ubiquitin-like) locus. First described by Dassa et al. [17], the T.th-BUBL1 locus (Fig. 1) consists of five UBLs, two Bacterial-Intein Like (BIL) elements and an ADP-ribosyl transferase (ART) domain, similar to bacterial toxins. Although BUBL1 is translated as a single protein, BIL-operated cleavage of the polypeptide chain has been hypothesized as a post-translational modification, likely responsible for the maturation and possible conjugation of the flanking UBLs [17] [2]. Such cleavage is the result of the side reaction of BIL splicing activity. In fact, while inteins accomplish splicing by efficiently linking the two flanking regions, BIL domains often induce cleavage of either or both the C- and the N-flanks [18] [19].

The two BILs in the BUBL1 locus act upstream of T.th-ubl5 and as a result T.th-ubl5 C-terminus remains permanently connected to the rest of the protein. Therefore, T.th-ubl5 can be considered as an integral part of a protein containing an ART domain. To date, the BUBL loci from T. thermophila and Paramecium tetraurelia represent the only co-occurrences of an UBL and an ART domain in the same protein. Interestingly, T.th-ubl5 lacks the hydrophobic patch responsible for the interactions with the UIM motif, precluding its recognition by most of ubiquitin partners. By structurally and sequentially comparing T.th-ubl5 with other integral UBLs, a novel biological function for the domain, with respect to the host protein, is proposed and demonstrated.

2. Materials and methods

2.1. Protein expression and purification

The coding sequence for T.th-ubl5, corresponding to residues 628–707 of the original locus (UNIPROT_P23659), was inserted between HindIII and BamHI sites of pHYRSF53 plasmid (gift from Dr. Hideo Iwai, Addgene plasmid # 64696) resulting in an N-terminally hexa-histidine tagged Smt3 fusion protein. Uniformly 15N, 13C, double-labelled T.th-ubl5 was produced in E. coli ER2566 strain (New England Biolabs), in 2-L M9 medium supplemented with 15NH4Cl and 13C6-glucose as sole sources of nitrogen and carbon, respectively. When OD600 reached 0.6, protein expression was induced by addition of a final concentration of 1 mM IPTG. The cells were grown for additional 4 h at 37°C before being harvested and centrifuged. Cells were then resuspended in Buffer A (300 mM NaCl, 50 mM sodium phosphate, pH 8.0) and lysed using Emulsiflex C3 homogenizer. After centrifugation at 37,000 rpm for 1 h at 4°C and removal of the cell debris, the supernatant was loaded onto a Ni-NTA column (GE Healthcare) and the protein was purified with a linear gradient of Buffer B (300 mM NaCl, 50 mM sodium phosphate, 250 mM imidazole, pH 8.0). The purified fusion-protein was digested with Up11 and the tag removed by a second IMAC purification. The pure T.th-ubl5 was collected in the flow-through.

For the splicing-hampered precursor, protein sequence comprising ubl4, BIL2 and ubl5 of the BUBL1 locus bearing alanine mutations at residues C77 and N219 (first and last residues of BIL2, respectively) was ordered as synthetic gene cloned in PET28b vector between Ncol and Xhol restriction sites. BL21(DE3) transformed E. coli cells were grown in LB/N medium [20] up to OD = 0.9, then transferred to 47°C incubator and induced with 1 mM IPTG concentration. After 20 min of heat-shock treatment the cells were cooled down to 20°C and left incubate O/N. Cells were harvested and centrifuged, and the pellet was resuspended with Buffer B (20 mM Hepes, 100 mM NaCl, pH 8) and lysed by sonication. The lysate was centrifuged at 23,000g for 1 h at 4°C, the supernatant was loaded onto a Ni-NTA column (GE Healthcare) and the protein was eluted with Buffer C (20 mM Hepes, 100 mM NaCl, pH 8, 0.5 M imidazole).

T.thRas gene was cloned into pGEX-4 T-1 vector between BamHII and Xhol restriction sites resulting in an N-terminally GST tagged fusion protein bearing an additional His tag at the C-terminus. Differently form the GST tag, the C-terminal His tag was not cleavable. The production of T.thRas as double-tagged recombinant protein was meant to maximize the purification efficiency. Transformed cells were grown up to OD = 0.6 and induced with 1 mM IPTG concentration for 3 h at 37°C. Harvested cells were then resuspended with Buffer B (20 mM Hepes, 100 mM NaCl, pH 8), lysed by sonication for 30 min and centrifuged at 23,000g for 1 h at 4°C. Cells debris was removed and the supernatant was loaded onto a GST-trap column (GE Healthcare). Protein was eluted with Buffer D (20 mM Hepes, 100 mM NaCl, pH 8, 20 mM glutathione).

2.2. NMR spectroscopy

15N, 13C-double labelled protein was dialyzed against 20 mM sodium phosphate buffer, pH 6 and concentrated using an Amicon ultra 4 concentrator (Millipore) to a final protein concentration of 4 mM. 10% D2O was added to the NMR sample for a total volume of 250 μL in a Shigemi tube. Spectra were acquired at 298 K on 850 MHz AVANCE III HD and 600 MHz AVANCE III spectrometers, both equipped with a triple-resonance cryoprobe. The sequential backbone assignment was carried out using [13C,15N]-HSQC, HNCA, HNCACB and CBCA(CO)NH experiments. Assignment of 1H and 13C side chain resonances was based on [13C,15N]-HSQC, (HCCO)NH, HCC=COSY and 15N resolved [1H,15N]-TOCSY. Distance restraints were derived from 13C- and 15N- edited NOESY-HSQC spectra using mixing
times of 75 ms and 80 ms, respectively. All spectra were processed with TOPSPIN 3.1 (Bruker Inc.) and analysed with CcPnMr [21].

2.3. NMR structure calculation

Structure calculation was performed using CYANA 3.0 software [22]. Two hundred conformers were generated based on automated NOE cross peak assignment [23]. The final number of NMR-derived restraints was 1321 of which 1196 NOE distances and 146 backbone angle restraints predicted with TALOS-N [24] [25]. The 20 lowest-energy structures were selected and energy minimized in explicit water using AMBER 14 [26]. The final structure was then analysed and validated using Protein Structure Validation Suite 1.5 (PSVS) [27]. Coordinates have been deposited in the Protein Data Bank under the accession code 5N9V and Chemical Shift Data Bank with the accession number 34106. Chemical Shift Resonance Data Bank with the accession number 34106. Chemical Shift Index (CSI) was calculated by the program CSI 3.0 [28].

2.4. Molecular dynamics simulations

In order to test the possibility of an interaction between T.th-ubl5 and the proteasome, we carried out atomistic MD simulations of T.th-ubl5 NMR solution structure in complex with the I-TASSER [29] structural predictions of the UIM1 and UIM2 motifs of the proteasome component Rpn10 from T. thermophila (UNIPROT_I7M6L0). T.th-UIM motifs were found through a BLAST search performed over the genome of T. thermophila, using H.sap-Rpn10 as query (Fig. S1). The following systems were set up: 1) H.sap-ubq/H.sap-UIM1, 2) H.sap-ubq/H.sap-UIM1-mut, 3) H.sap-ubq/T.th-UIM1, 4) T.th-ubl5/H.sap-UIM1, 5) T.th-ubl5/T.th-UIM1, 6) T.th-ubl5-mut/T.th-UIM1 7) T.th-ubl5/H.sap-UIM2 8) T.th-ubl5/T.th-UIM2 9) T.th-ubl5-mut/T.th-UIM2 10) H.sap-ubq/H.sap-UIM2. Starting from system 1 (Fig. S2), showing the native structural insights of the known interaction (PDB ID: 1YX5), key residues were mutated (Fig. 3) and domains progressively exchanged to form hybrid complexes. In particular, the systems (n’ 5 and 8) were constructed by domain-domain superimposition of T.th-ubl5 and T.th-UIM motifs to the experimental structures of human homologs ubq/Rpn10-UIM1 (1YX5) and ubq/Rpn10-UIM2 (2KDE), respectively. Seven additional hybrid systems were constructed by either pairing one domain with its partner’s ortholog (systems 3, 4, 7) or by introducing species-specific mutations resulting in chimeric domains (2, 6, 9, 10) (Fig. 3). Such systems were simulated for comparative interaction analysis. All amino acids were considered in their standard protonation states, that is K and R protonated, D and E deprotonated and H neutral with delta or epsilon nitrogen protonated. The systems were solvated with TIP3P [30] water molecules in a cubic box, whose size was defined by a 2 nm distance between the box and the solute. Sodium and chloride ions were added to simulate a 0.1 M salt concentration while maintaining electrical neutrality. Verlet integrator was used with a non-bonded interaction cut-off of 12 Å along with a 2 fs time step. PME [31] was used to treat the long-range electrostatics. For each system, two simulation replicas were made, which differed in equilibration steps (1.5 and 2.5 ns), thus different starting structures for subsequent production MD. About 1 μs long production runs were performed at constant temperature and pressure of 310 K and 1 atm, respectively. The Parrinello-Rahman barostat and Nose-Hoover thermostat implemented in GROMACS 5.1.4 [32] were used for the purpose. The entire model system comprising protein, water and ions was based on CHARMM36 force field [33]. Simulation trajectory analysis was carried out with VMD [34] and GROMACS.

Contacts between T.th-ubl5/H.sap-ubq and UIM were calculated over each simulation trajectory by using a distance cut-off of 3 Å. The following were considered as active residues for UIM motifs: I-A-L-A-L (10–14) for H.sap-UIM1, I-A-Q-A-M (10–14) for T.th-UIM1 and H.sap-UIM1-mut, I-A-Y-A-M (10–14) for H.sap-UIM2 and M-N-Q-A-I (10–14) for T.th-UIM2 and H.sap-UIM2-mut. In addition to the three “core residues” of the hydrophobic patch, six other were selected as active residues for ubq/UbL domains because of their solvent-exposure and proximity to the patch. Overall, the chosen residues were: K6, L8, R42, I44, A46, T66, H68, V70, R72 for H.sap-ubq, E6, Q8, S42, K44, A46, E69, H71, E73, K75 for T.th-ubl5 and residues E6, L8, R42, I44, A46, E69, H71, V73, K75 for T.th-ubl5-mut.

Data analysis was carried out by implementing empirical parameters which were used as indicators of the interaction strength and dissociation rate. The first one, named LCF (Loss of Contact per Frame) was calculated as:

\[
LCF = \frac{\sum_{d=1}^{nf} (C_{max} - C_d)}{nf}
\]

where \(nf\) corresponds to the total number of frames (8334), \(C_{max}\) corresponds to the highest number of contacts observed in the simulation (always equivalent to those observed in the first frame) and \(C_d\) is the number of contacts of a specific frame. Average LCF was then calculated as the mean value of the %FNB values from two simulation replicas. As no specific number of contacts could be used as threshold to
differentiate bound from unbound state, and %FNB not being informative of the association rate of the domains, the FNB (Frames Not Bound) parameter was also calculated as the percentage of frames with < 1 contact. Average FNB value was also calculated as mean value between the two replicas.

2.5. Sequence alignment

ECOD [35] and UbSRD [36] databases were used for the identification of integral UbL structures. Domain preselection was carried out by collecting entries classified as UbL in UbSRD together with non-SUMO-like structures retrieved by ECOD with an RMSD < 3 Å to the T.th-ubl5 NMR structure (lowest-energy conformer submitted). False-positive structures, showing protein-ubiquitin interactions were discarded as well as duplicate structures. In the latter case, if any, structures of UbL together with the host protein were chosen over the isolated UbL. Structures of the same protein from different organisms were considered different and no selection was made. Each identified UbL was then verified to be integral to a host protein by checking its sequence in the respective Uniprot page from the PDB.

In the end, 79 structures of integral UbLs were retrieved and used for comparison. Multiple sequence alignment was carried out with MAFFT [37]. As structural information has been shown to improve the accuracy of the results [38] [37], the alignment was corrected with the MAFFT structural-alignment option.

The same procedure was used for structural alignment of T.th-ubl5 with FERM domains (Fig. 4). Structures were selected by a DALI search [39] of T.th-ubl5.

2.6. SPR experiments

SPR experiments were carried out using a SensiQ Pioneer system. Immobilization of the ligand (a mutated ubl4-BIL2-ubl5 splicing precursor bearing C77A and N219A alanine mutations at N- and C-termini of BIL2) was carried out essentially as in Genovese et al. [46]. The sensor chip (COOH1) was chemically activated by a 100 μl injection of a 1:1 mixture of N-ethyl-N’-3-(diethylaminopropyl)carbodiimide (200 mM) and N-hydroxysuccinimide (50 mM) at a flow rate of 5 μl/min. The ligand was immobilized on activated sensor chips via amine coupling. The immobilization was carried out in 20 mM sodium acetate at pH 4.0; the remaining unreacted groups were blocked by injecting 1 M ethanolamine hydrochloride (100 μl). The amount of immobilized ligand was detected by mass concentration-dependent changes in the refractive index on the sensor chip surface, and corresponded to about 200 resonance units (RU); a flow cell both activated and deactivated in the same conditions, but without immobilized ligand, was used as a reference. Analytes, i.e. GST and GST-ubl4-BIL2-ubl5 splicing precursor in HSP buffer (10 mM Hepes pH 7.4, 150 mM NaCl, 0.005% surfactant P20), was injected on the sensor chip at a constant flow (30 μl/min). A FastStep procedure was used: the analytes were automatically diluted in HSP and injected by 6 serial doubling steps (step contact time = 20 s, nominal flow rate = 100 μl/min). At the following time points: 1) 0–30 s; 2) 31–60 s; 3) 61–90 s; 4) 91–120 s; 5) 121–150 s; 6) 151–162 s, analyte concentrations were: 1) 0.0312 μM; 2) 0.0625 μM; 3) 0.125 μM; 4) 0.25 μM; 5) 0.5 μM; 6) 1 μM (sensorgams A) or 1) 0.0937 μM; 2) 0.1875 μM; 3) 0.375 μM; 4) 0.75 μM; 5) 1.5 μM; 6) 3 μM (sensorgams B). The increase in RU relative to baseline indicates complex formation, whereas the decrease in RU after 162 s represents dissociation of analytes from immobilized ligands after injection of buffer HSP. The sensorgams were analysed using the SensiQ Qdat 4.0 program, using full fittings with 1 site.

3. Results

3.1. T.th-ubl5 solution structure

The [15N,1H]-HSQC spectrum of T.th-ubl5 displays well-spread resonances (Fig. 2A), indicating a properly-folded structure. A nearly complete assignment of backbone (99.7%) and side chain (97.4%) resonances was obtained. The backbone superposition of the final 20-structure bundle (Fig. 2B) shows a well-defined fold belonging to the beta-grasp, with an average atomic root mean square deviation (RMSD) to the mean structure of 0.52 ± 0.08 Å for ordered regions, defined by residues 3-65 and 70-77. Additional structural statistics are provided in Supplementary Table S1. As expected for a ubiquitin-like domain, T.th-ubl5 secondary structure consists of an alpha helix lying on a mixed, five-stranded beta sheet with a 2-1-5-3-4 order (Fig. 2C). As predicted from the backbone chemical shift index (Fig. 2A, inset), additional single-turn helices occur in the regions connecting the alpha helix to the third beta strand, and the fourth and fifth beta strands.

3.2. Molecular dynamics studies on the interaction between T.th-ubl5 and proteasome

Although excellent conservation of the β-grasp fold by T.th-ubl5 highlights no unexpected structural features, T.th-ubl5 differs from common ubiquitin and UbL domains by lacking the characteristic hydrophobic interaction surface mainly responsible for proteasomal recognition. In case T.th-ubl5 is able to interact with the proteasome, the exerted function might be that of either fostering the host protein degradation or recruiting the ART domain onto the proteasome, probably regulating some of its subunits. In order to describe how the lack of the hydrophobic patch can be detrimental for the interaction with the proteasome, several model systems were constructed and atomistic MD simulations were performed.

Simulation system 1 (Fig. S2) shows a stable complex between H.sap-ubq and H.sap-UIM1, with a loss of < 1 contact per frame and only about 2% of the whole simulation time showing complete dissociation (Table 1). When H.sap-UIM1 was mutated with T.thermophila five-residue stretch (system 2), a destabilizing effect was observed with
two and four times increase in LCF and %LFB, respectively. *H.*sap-ubq and *T.*th-ubl5 were then simulated in complex with their reciprocal UIM1 partners (systems 3–4). While system 3 is comparable with system 1, due to the good conservation of the interacting residues between the two ortholog UIM1 domains, system 4 shows a drastic increase in the dissociation rate and reduction of the interaction strength, though with some variations in the two simulation replicas. The strong correlation between systems 1 and 3 (in contrast to setup 4) suggests that the interaction tolerates modifications on the UIM1 domains much better than on the UbL.

As hypothesized above, system 5, comprising *T.*th-ubl5 in complex with *T.*th-UIM1, showed weaker interaction strength, the very same dissociation rate and reduction of the interaction strength, though with some variations in the two simulation replicas. The strong correlation between systems 1 and 3 (in contrast to setup 4) suggests that the interaction tolerates modifications on the UIM1 domains much better than on the UbL.

As hypothesized above, system 5, comprising *T.*th-ubl5 in complex with *T.*th-UIM1, showed weaker interaction strength, the very same dissociation rate and reduction of the interaction strength, though with some variations in the two simulation replicas. The strong correlation between systems 1 and 3 (in contrast to setup 4) suggests that the interaction tolerates modifications on the UIM1 domains much better than on the UbL.

Interaction with UIM2 was also investigated through systems 7, 8, 9, and 10, analogous to systems 4, 5, 6, and 2, respectively. A similar variation between simulation replicas as in system 4 is observed in setup 7, nevertheless, as expected, system 7 (replica ‘a’) shows a low stability, up to 3 to 4 times lower than that of system 4, hence supporting the detrimental effect on the interaction due to the lack of hydrophobic residues on *T.*th-ubl5. Systems 6 and 9 also show a very similar amount of maximum contacts, although the “rescuing effect” provided by the humanized *T.*th-ubl5 in system 9 was not as high as the one observed for UIM1 in system 6. A great resemblance is observed between systems 2 and 10, where all parameters indicate an on-off association based on conserved interactions provided by the hydrophobic patch on *H.*sap-ubq, therefore providing good control for our simulations.

Finally, system 8 shows surprisingly low LCF and %FNB values suggesting an interaction similar, if not stronger than that in system 1. Interestingly, the maximum number of native contacts found over the two trajectories is 2, which implies that LCF and %FNB cannot be used to derive any reliable information on the actual interaction state. An alternative explanation for this behaviour may be the presence of a secondary interaction surface, which does not include the active residues used for native contacts calculation. A close look to the trajectory reveals in fact, that *T.*th-UIM2 docks itself on the N-terminus side of the *T.*th-ubl5 barrel. A native contacts calculation performed between the *T.*th-UIM2 five-residue stretch and the whole *T.*th-ubl5 domain shows that no additional contacts take place demonstrating that the interaction responsible for the unexpected LCF and %FNB values is unspecific i.e. does not occur between any conserved patches of the two partners. Moreover, as the N-terminus side of the *T.*th-ubl5 barrel would not be accessible to *T.*th-UIM2...
due to the steric hindrance by BIL2, such predicted interaction cannot 
take place in vivo. According to this, this secondary interaction is not 
detected in system 9 where the residues in question remain unaltered. 
Overall, atomistic simulations show that the presence of T.th-ubl5 
cripples the stability of the interaction because the lack of the hydrophobic 
patch weakens or prevents the association with the interaction cleft.

3.3. Comparison with other integral UbLs

In order to investigate the possible biological function of T.th-ubl5, 
we performed a multiple sequence alignment with other integral UbLs 
of known structure. Seventy-eight sequences were aligned with those of 
T.th-ubl5 and ubiquitin using MAFFT [37] (Fig. 4) (see Methods).
Integral UbLs sharing a similar sequence were automatically clustered together in different subgroups. Members of each subgroup were found to be integral to a specific type of protein, indicating how integral UbLs have evolved to improve the structural relationship with their host proteins (Fig. 4). Because T.th-ubl5 does not feature the conserved hydrophobic patch, we tried to identify which other integral UbLs lack this primary interaction surface. Out of the 78 sequences, 74 were found to lack highly hydrophobic residues (I, L, F, C, defined according to their hydropathy index [41]) in at least one of the three expected positions (residues 8, 44, 70, following the ubiquitin numbering). Only 27 of 74 UbL structures included a host protein or a part of it. Interestingly, almost half of these 27 UbL structures were found to be FERM domains. The FERM domains, such as ezrin, talin, moesin, merlin and radixin together with some membrane associated kinases are...
membrane-localization modules found in several proteins associated with the cytoskeleton [42] [43]. FERMs are composed of three subdomains called F1, F2 and F3, the first of which is the UbL. Sequence analysis has identified the F1 subunit as a RA (Ras-associated) domain [44]. Experimental validation of the actual interaction between FERM and Ras was subsequently reported [45] [46].

In order to determine whether T.th-ubl5 could also be a RA domain, we performed a new multiple sequence alignment with FERM domains of known structure retrieved from DALI server using T.th-ubl5 as input structure. Twelve non-redundant structures of UbLs from FERM domains were identified and aligned to T.th-ubl5 as above (Fig. 5). Interestingly, the $\beta_2$-strand (residues 10–17) is rather conserved in terms of residue type composition, with 4–5 hydrophobic residues and 3–4 hydrophilic ones, sequentially arranged with alternative polarity. Hydrophilic residues in positions 13 and 15 are conserved. An important sequence conservation for the same set of residues can be observed between T.th-ubl5 and the FERM domain of SNX17 (PDB ID: 4GXB) for which the interaction with Ras has been experimentally determined [47]. In particular, G10, K12, N16 and V17 are conserved between the two. Residues in position 11, 13 and 15 also belong to the same amino acid type. The only non-conserved residue is K14 which in SNX17 is a valine. Another conserved residue between T.th-ubl5 and SNX17 is K33 at the C-terminus of the main alpha helix (Fig. 5). In KRIT1 FERM domain, the role of the arginine in this position was proved to be essential for the interaction with the Ras protein Rap1 [46].

In order to further investigate the putative Ras-binding epitope of T.th-ubl5, we structurally superimposed T.th-ubl5 with the SNX17 FERM domain and calculated the backbone RMSD between the two domains (Fig. 6). The N-terminus of the domain encompassing the Ras-binding epitope showed an RMSD value of ca. 1.93 Å, indicating an overall structural similarity, in contrast with the C-terminus. Similarly, when sequences of T.th-ubl5 and SNX17 FERM domains were aligned,
showed a strong and specific binding between BUBL1 and GST-Ras concentrations. Ras interacts strongly with ubl4-BIL2-ubl5: sensorgrams T.thermophila immobilized via amino-coupling on to the COOH1 sensor chip while mutations at N- and C-termini of BIL2 (C77A and N219A) was then conjugated. A mutated splicing precursor bearing splicing hampering sequence (RGG) are well conserved and ubl4 holds the possibility to be not feature the Ras binding epitope. Instead, residues forming the hydrophobic surface responsible for ubiquitin association and recognition, this was used as a further criterion to functionally distinguish the different domains. Interestingly, the Ras-binding domain of FERM was recurrently identified as the host protein of integral UbLs lacking the hydrophobic patch. FERM-containing proteins have been shown to play an important role in the regulation of endocytosis and vesicle trafficking like for example, sorting nexins SNX17, SNX27 and SNX31 have been found to coordinate the spatio-temporal organization of endosomes by associating with phosphoinositides and multiple proteins of the endosomal regulating machinery [47]. Alignment of T.th-ubl5 and FERM UbLs revealed a significant sequence (and also structural) conservation with the SNX17 integral UbL for which the Ras-binding function has been established. Interaction of T.th-ubl5 with Ras GTPase proteins was therefore investigated by SPR experiments, which revealed that T.th-ubl5 is able to bind Ras with sub-micromolar affinity. Even though the actual length of the protein resulting from the processing of the BUBL precursor is un-predictable, we have shown the role of T.th-ubl5 in what is the minimal host protein environment.

3.4. T.th-ubl5/T.th-Ras interaction

Given the sequence conservation of the Ras binding epitope between T.th-ubl5 and SNX17 F1 domain, we used the sequence of the SNX17 binding partner H-Ras to identify the most similar Ras protein in T. thermophila. From the analysis of available structures in PDB (i.e. 4HD0, 6AMB), the interaction of Ras proteins with the ubiquitin fold relies on the conserved stretch of amino acids E-D-S-Y-R forming the second beta strand of the protein. A BLAST search over the proteome of T.thermophila SNX17 binding partner H-Ras to identify the most similar Ras protein in T.thermophila revealed to have sequence similarity with the splicing processing. In fact, the interaction of Ras, SPR experiments were carried out. As T.th-ubl5 is simultaneously part of the ART domain and of the "BIL2 splicing system", we were particularly interested in testing the binding of Ras to T.th-ubl5 having both ubl4 and BIL2 at its N-terminal, so as to investigate if its occurrence can occur before and/or concomitantly with the splicing process. In fact, the interaction of T.th-Ras with T.th-ubl5 alone would not be informative on whether it can take place before the protein splicing reaction. Contrarily to T.th-ubl5, ubl4 does not feature the Ras binding epitope. Instead, residues forming the hydrophobic patch as well as the C-terminal conjugation consensus sequence (RGG) are well conserved and ubl4 holds the possibility to be conjugated. A mutated splicing precursor bearing splicing hampering mutations at N- and C-termini of BIL2 (C77A and N219A) was then immobilized via amino-coupling on to the COOH1 sensor chip while GST (Fig. S5) and GST-T.th-Ras (Fig. S6) were injected at different concentrations. Ras interacts strongly with ubl4-BIL2-ubl5: sensorgrams showed a strong and specific binding between BUBL1 and GST-Ras (with submicromolar KD values), with very slow observed dissociation, while almost no interaction takes place between BUBL1 and GST (Fig. S5), at the same concentrations used for Tth-GST-Ras. Fig. S6 reports the difference between GST-T.th-Ras and GST in the same experimental conditions; control experiments performed in the presence of 50 mM imidazole (Fig. S7), and control experiments performed by using GST-T.th-Ras as ligand and mutated ubl4-BIL2-ubl5 splicing precursor as analyte (Fig. S8), show a similar behavior, with submicromolar KD values (K0 = 500 nM ± 200 nM) and with very slow observed dissociation.

4. Discussion

We have presented the NMR solution structure of a ubiquitin-like domain integral to an ADP-ribosyltransferase protein. Because of the low sequence similarity among integral UbLs and the lack of information about the structure of the host protein, we first investigated whether T.th-ubl5 can interact with the proteasome, hence carrying out protein-degradation functions. Results from molecular dynamic simulations of several interaction models based on available deposited structures ruled out this possibility. The possible function of T.th-ubl5 was then studied by comparing structural features with other integral UbLs available in the PDB selected by using co-translational association with the host protein as the filtering parameter. As T.th-ubl5 lacks the residues forming the hydrophobic surface responsible for ubiquitin association and recognition, this was used as a further criterion to functionally distinguish the different domains. Interestingly, the Ras-binding domain of FERM was recurrently identified as the host protein of integral UbLs lacking the hydrophobic patch. FERM-containing proteins have been shown to play an important role in the regulation of endocytosis and vesicle trafficking like for example, sorting nexins SNX17, SNX27 and SNX31 have been found to coordinate the spatio-temporal organization of endosomes by associating with phosphoinositides and multiple proteins of the endosomal regulating machinery [47]. Alignment of T.th-ubl5 and FERM UbLs revealed a significant sequence (and also structural) conservation with the SNX17 integral UbL for which the Ras-binding function has been established. Interaction of T.th-ubl5 with Ras GTPase proteins was therefore investigated by SPR experiments, which revealed that T.th-ubl5 is able to bind Ras with sub-micromolar affinity. Even though the actual length of the protein resulting from the processing of the BUBL precursor is unpredictable, we have shown the role of T.th-ubl5 in what is the minimal host protein environment.

By demonstrating that T.th-ubl5 is able to decoy Ras GTPase protein of T. thermophila, we speculate that this interaction could promote the ADP-ribosylation of Ras by the host protein as well as its ubiquitination by BIL2. As Ras mono-ADP-ribosylation has been observed to be performed only by pathogen effectors against host GTPases [48] [49] [50], the putative ADP-ribosylation of T.th-Ras carried out by the host BUBL locus would open an interesting new field of investigation, introducing a novel type of regulation of trafficking and signalling pathways. In particular, the complex combination of ubiquitin and ADP-ribosylation resulting in a novel regulative signal for GTPases has already been described by Qiu et al. [50]. The study showed that in Legionella pneumophila the ADP-ribosylation activity of effector proteins can control the maturation of the pathogen-containing endosomes by hijacking the host's ubiquitin system and use it against Rab GTPases. Interestingly, T. thermophila has also been found to be an occasional host of L. pneumophila [51].
capacity to recruit other proteins, suggest that it could be an important platform of protein modification and sorting. Unlike most intein-containing proteins, BIL exteins are independent domains which maintain their functions regardless of the presence of BIL. In this case, unlike many essential proteins, the role of the splicing element activity is not to regulate the host protein merely through a switch-on switch-off mechanism, but provides a finer degree of regulation by adding complexity to the whole system. In particular, BIL2 is flanked by two domains with different functions, being ub4 a conjugatable moiety and T. th-ub5 the decy recursing Royal on to a larger ADP-ribosylation domain. Moreover, the strategical position of ub4 at the N-terminus of BIL2 links its C-terminal conjugation motif R-G-G to the thioester-forming cysteine of the splicing domain. This strongly emulates the ubiquitination intermediate where ubiquitin is attached to the E3 ubiquitin ligase before its conjugation to the final target. In light of these observations, we speculate that BUBL1 might act as an E1-E2-E3-free ubiquitination platform onto which the target proteins are lured in order to undergo post-translational modifications. If this was true, it would furthermore constitute a remarkable example of intein’s molecular parsarism turning into beneficial partnership between the host and the protein splicing element, with considerable impact from an evolutionary point of view.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgements

This work was supported by the Integrative Life Science Doctoral Program (ILS). VS acknowledges funding from the Academy of Finland, University of Helsinki, Sigrid Juselius Foundation and the Magnus Ehrnrooth Foundation. CSC-IT Center for Science, Finland, is acknowledged for computational resources. The authors wish to thank Dr. Joseph Harrison for providing the updated list of ubiquitin-like domains present in UbSRD database. We acknowledge “Quality methods for Design of Experiments in Scientific Research”, in the FaReBiodiQualità Program (ILS). VS acknowledges funding from the Academy of Finland, and mechanism, Chem. Rev. (2017), https://doi.org/10.1021/acs.chemrev.6b00737. Evaporative point of view.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgements

This work was supported by the Integrative Life Science Doctoral Program (ILS). VS acknowledges funding from the Academy of Finland, University of Helsinki, Sigrid Juselius Foundation and the Magnus Ehrnrooth Foundation. CSC-IT Center for Science, Finland, is acknowledged for computational resources. The authors wish to thank Dr. Joseph Harrison for providing the updated list of ubiquitin-like domains present in UbSRD database. We acknowledge “Quality methods for Design of Experiments in Scientific Research”, in the FaReBiodiQualità Program (ILS). VS acknowledges funding from the Academy of Finland, and mechanism, Chem. Rev. (2017), https://doi.org/10.1021/acs.chemrev.6b00737.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgements

This work was supported by the Integrative Life Science Doctoral Program (ILS). VS acknowledges funding from the Academy of Finland, University of Helsinki, Sigrid Juselius Foundation and the Magnus Ehrnrooth Foundation. CSC-IT Center for Science, Finland, is acknowledged for computational resources. The authors wish to thank Dr. Joseph Harrison for providing the updated list of ubiquitin-like domains present in UbSRD database. We acknowledge “Quality methods for Design of Experiments in Scientific Research”, in the FaReBiodiQualità Program (ILS). VS acknowledges funding from the Academy of Finland, and mechanism, Chem. Rev. (2017), https://doi.org/10.1021/acs.chemrev.6b00737.


