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

Valerio Chiarini[a,*], Helena Tossavainen[b], Vivek Sharma[a,c], Gianni Colotti[d,**]

[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 Jyvaskyla, P.O. Box 35, FI-40014 Jyvaskyla, 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*

ARTICLE INFO

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

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, I44 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 *β*-grasp fold are better tolerated. It is reasonable to think that, in this case, ubiquitin fold operates just as a scaffold to display the hydrophobic patch driving protein-protein interaction.

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 (BIL-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_Q236S9), was inserted between *Hin*dIII 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 ^{15}N, ^{13}C- double labelled *T.th*-ubl5 was produced in *E. coli* ER2566 strain (New England Biolabs), in 2-L M9 medium supplemented with ^{15}NH$_4$Cl and ^{13}C$_6$-glucose as sole sources of nitrogen and carbon, respectively. When OD$_{600}$ 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 37,000g 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 Upl1 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 *Nco*I and *Xho*I restriction sites. BL21(DE3) transformed *E. coli* cells were grown in LBNB 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.th*Ras gene was cloned into pGEX-4 T-1 vector between BamHI and *Xho*I 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.th*Ras 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

^{15}N, ^{13}C-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% D$_2$O 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 [^{15}N,^{1}H]-HSQC, HNCA, HN(CO)CA, HN(CA)CO, HNCO, HNCACB and CBCA(CO)NH experiments. Assignment of ^{1}H and ^{13}C side chain resonances was based on [^{13}C,^{1}H]-HSQC, (H)CC(CO)NH, HCCH-COSY and ^{15}N resolved [^{1}H,^{1}H]-TOCSY. Distance restraints were derived from ^{13}C- and ^{15}N- edited NOESY-HSQC spectra using mixing
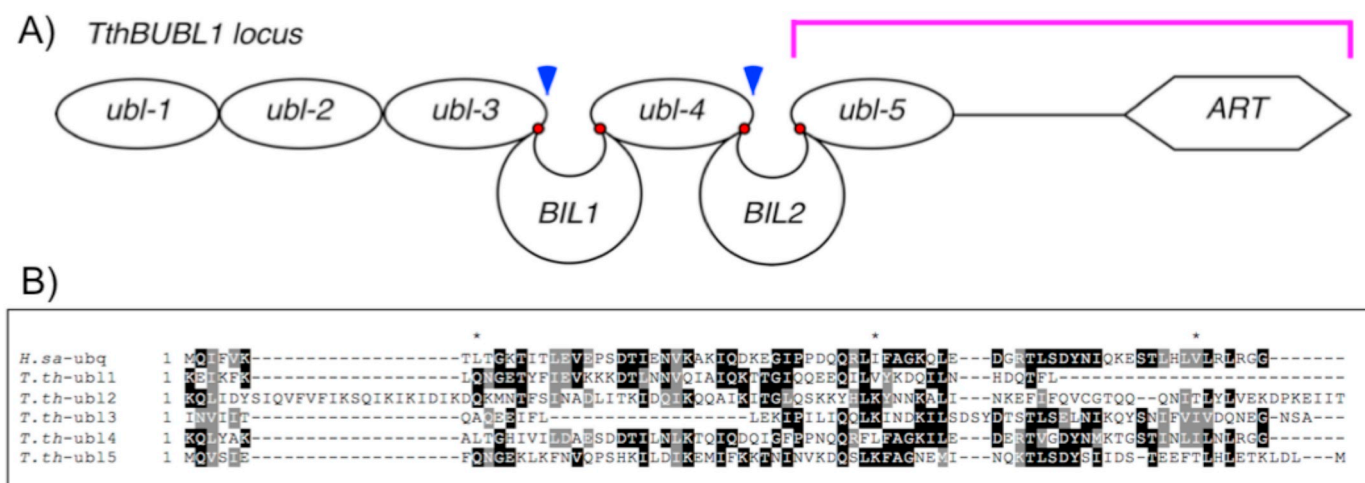
Fig. 1. A) Schematic depiction of *Tetrahymena thermophila* BUBL1 locus. BIL-cleaving sites are highlighted by red circles while blue arrows indicate UbL conjugation ends generated from the cleavage. Conjugation of ubl3 includes that of ubl1 and ubl2 as no other domains are present in between. The purple segment defines the minimal length of the ART-containing protein. B) Sequence alignment between ubiquitin and the five UbL domains of the BUBL locus. Asterisks indicates the hydrophobic patch core residues relative to ubiquitin sequence. Structural depiction of the residues conservation is shown in Fig. S3.

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 NOESY 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]. Atomic coordinates have been deposited in the Protein Data Bank under the accession code 5N9V and the chemical shifts have been deposited in the Biological Magnetic 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: L-A-L-A-L (10–14) for *H.sap*-UIM1, L-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, T69, H71, E73, K75 for *T.th*-ubl5 and residues E6, L8, R42, I44, A46, T69, 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:

$$LCF = \left[\sum_{n=1}^{nf} (C_{max} - C_{nx}) \right] / nf$$

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_{nx} 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 two 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. [40]. 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-*T.th*-Ras in HSP buffer (10 mM Hepes pH 7.4, 150 mM NaCl, 0.005% surfactant P20), were 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) 0–40 s; 2) 41–80 s; 3) 81–120 s; 4) 121–160 s; 5) 161–200 s; 6) 201–220 s, analyte concentrations were: 1) 0.0625 μM; 2) 0.125 μM; 3) 0.25 μM; 4) 0.5 μM; 5) 1 μM; 6) 2 μM (sensorgram A) or 1) 0.25 μM; 2) 0.5 μM; 3) 1 μM; 4) 2 μM; 5) 4 μM; 6) 8 μM (sensorgram B). The increase in RU relative to baseline indicates complex formation, whereas the decrease in RU after 220 s represents dissociation of analytes from immobilized ligands after injection of buffer HSP. The sensorgrams were analysed using the SensiQ Qdat 4.0 program, using full fittings with 1, 2 and 3 sites. The best curve fitting was obtained with 2 sites. The experiment was repeated in a new sensor chip, treated in the same fashion, with a level of immobilized ligand of about 150 resonance units (RU): the GST-*T.th*-Ras analyte was injected at the same concentration of the experiment in sensorgram B, in HSP buffer and in HSP buffer + 50 mM imidazole.

In another SPR experiment, 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 GST-*T.th*-Ras 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 25 resonance units (RU); a flow cell both activated and deactivated in the same conditions, but without immobilized ligand, was used as a reference. The analyte, i.e. mutated 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 (sensorgrams 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 (sensorgrams 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 sensorgrams 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 [^{15}N,^{1}H]-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
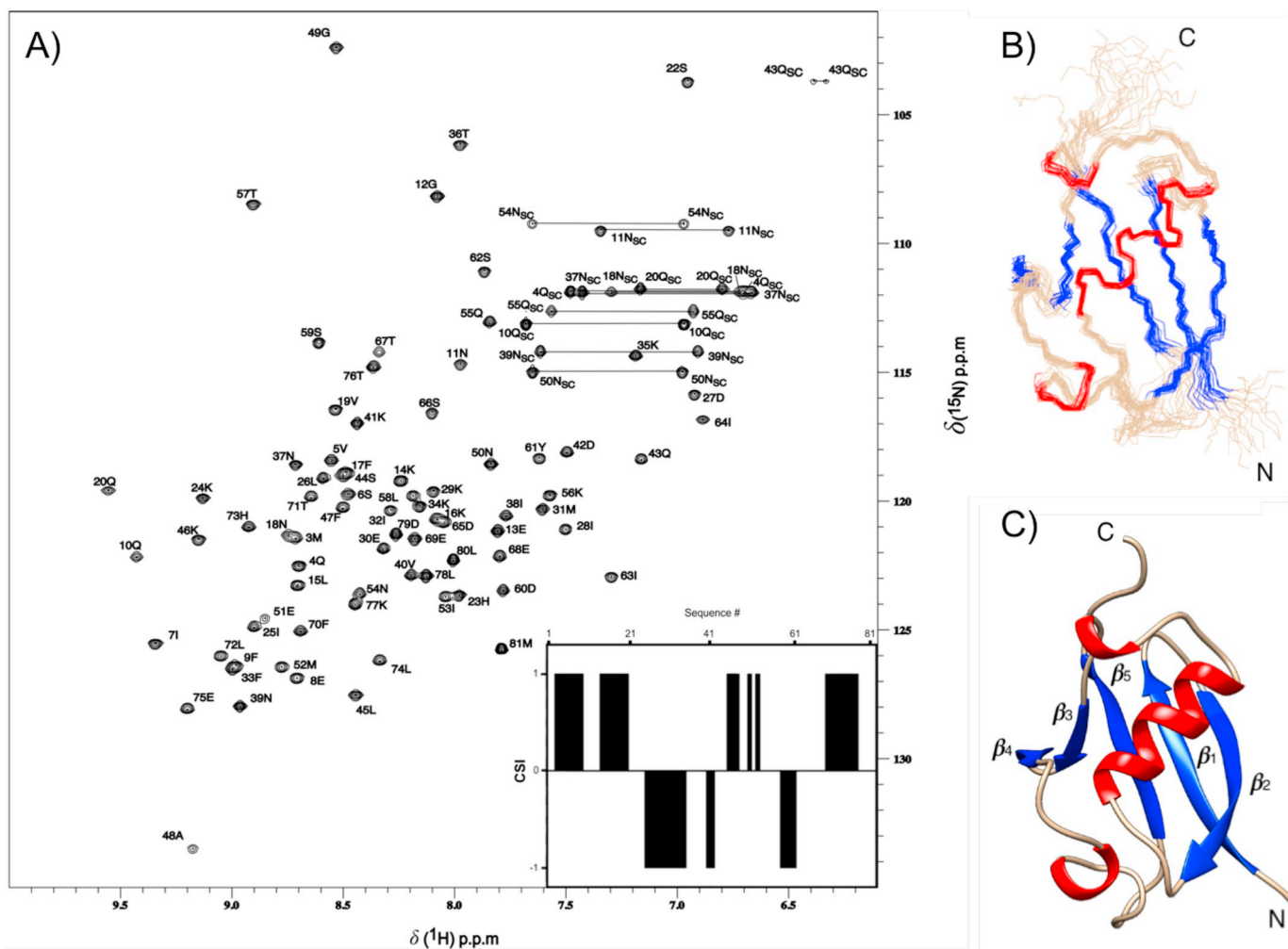
Fig. 2. A) [^{15}N,^{1}H]-HSQC spectrum of 4 mM *T.th*-ubl5 recorded at 600 MHz and 298 K. Asn and Gln side chain peaks are indicated (lines). Peaks from residues 1 and 2 (cloning artefacts) are not detectable. In the inset is presented the chemical shift index calculated with CSI 3.0 [28]. B) NMR solution structure of *T.th*-ubl5 (res 1–2 and 3–81 of the NMR assignment, the latter ones equivalent to aa 629–707 of the BUBL1 locus, UNIPROT_Q236S9), ensemble of 20 structures. C) Lowest-energy conformer. Secondary structure elements are shown in blue (beta sheets) and red (alpha helices), and the beta strand numbering is indicated. Figures were prepared with CHIMERA [52]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 of system 4, being over six times higher than that of system 3, indicating that *T.th*-ubl5 is unfit to bind the proteasome unlike the majority of UbLs. However, interestingly substitution of the interacting residues of *T.th*-ubl5 with human ones (system 6) considerably compensated the interaction with *T.th*-UIM1 and improved the stability of the complex.

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
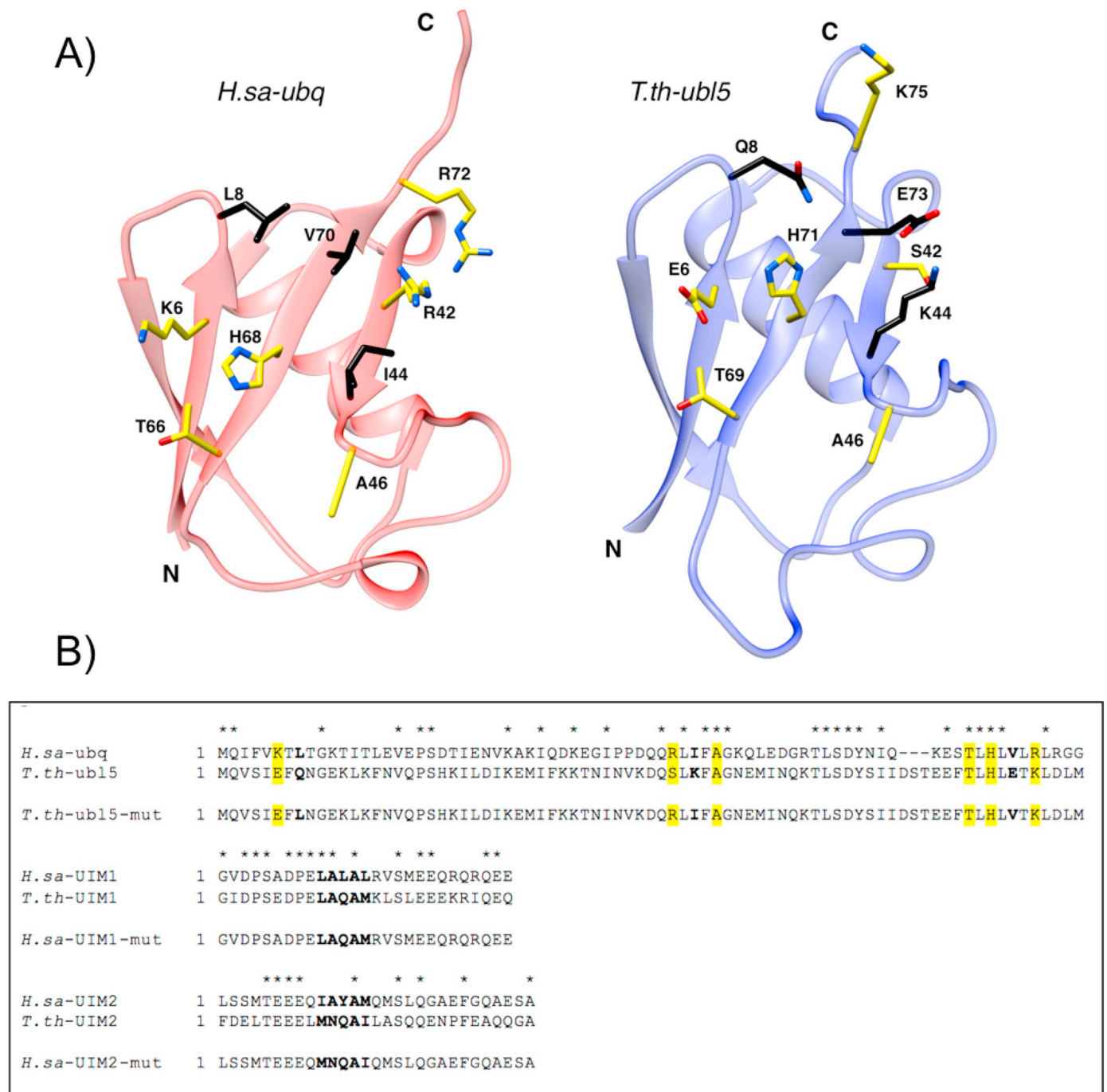
Fig. 3. A) Structural representation of *H.sa*-ubq (PDB ID: 1UBQ) and *T.th*-ubl5 (PDB ID: 5N9V), on the left and right, respectively. Residues involved in the formation of the main interaction surface are shown as sticks and labelled by one-letter codes. Hydrophobic-patch core residues are colored in black while proximal, solvent-exposed residues are shown in yellow. Residue numbering for *T.th*-ubl5 does not include the cloning artefacts (E6 corresponds to E8 in the deposited NMR assignments).
B) Sequences of different domains used in MD simulation systems: *H.sa*-ubq (PDB ID: 1UBQ), *T.th*-ubl5 (PDB ID: 5N9V), *H.sa*-UIM1 (PDB ID: 1YX5), *H.sa*-UIM2 (PDB ID: 2KDE) *T.th*-UIM1/*T.th*-UIM2 (UNIPROT_I7M6L0). For each ubq/UbL and Rpn10 domain the core residues of the "hydrophobic patch" and the UIM1/UIM2 motifs, respectively, are highlighted in bold. As in panel A, in yellow are indicated solvent-exposed residues which surround the hydrophobic patch on the beta-sheet side of the ubq/UbL domains and were considered as active residues for native-contacts calculations. Conserved residues between ortholog and chimeric domains are indicated with asterisks. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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).
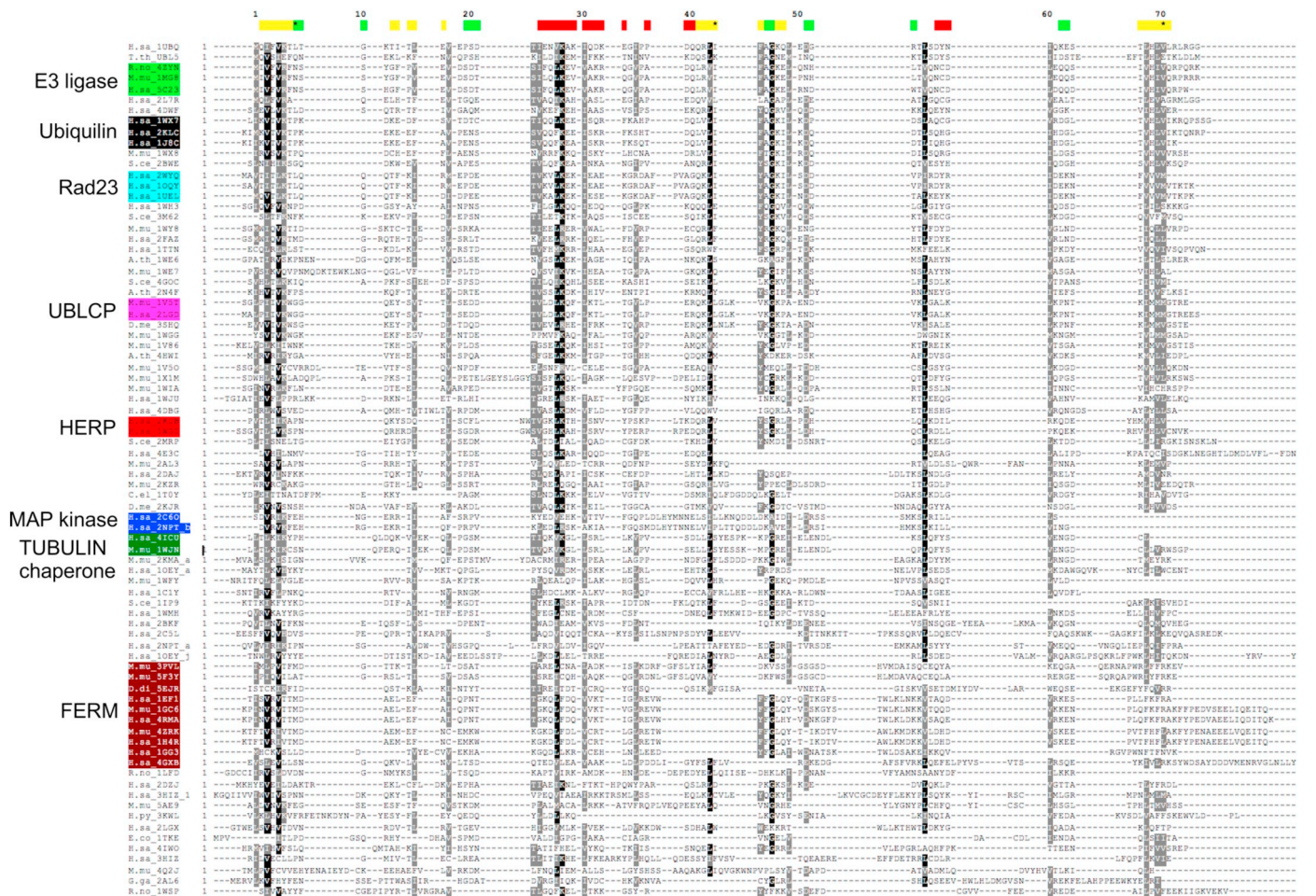
Fig. 4. Multiple sequence alignment of integral UbLs retrieved from UbSDR and ECOD databases, *T.th*-ubl5 and ubiquitin. Domains close in sequence were automatically grouped and are highlighted with different color codes. ULDs from the same group are integral to the host-protein indicated. On the top, residue numbering relative to ubiquitin sequence is shown. Secondary structure of ubiquitin is indicated with color-based code: yellow for beta strand, red for helices and green for turns. Asterisks mark the positions of residues forming the hydrophobic patch in ubiquitin. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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
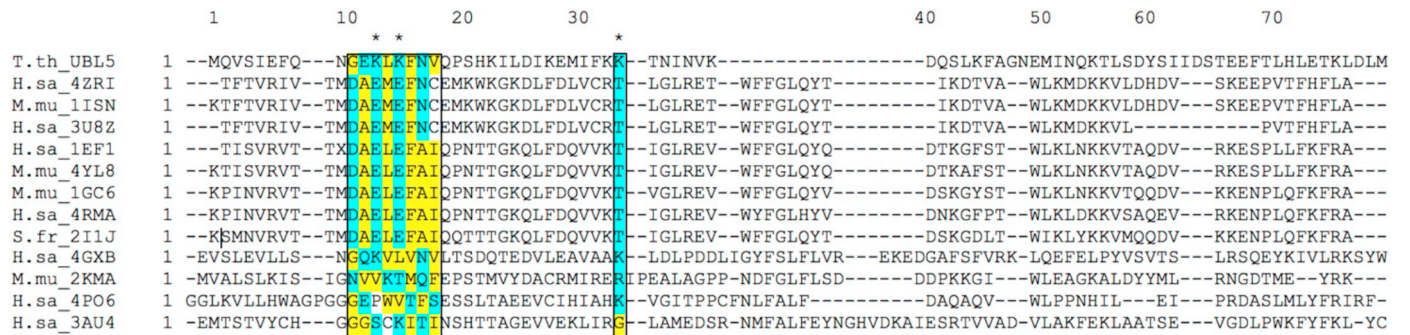

Fig. 5. Structural alignment of FERM-UbLs retrieved from DALI search of *T.th*-ubl5. Boxes frame the residues responsible for the interaction with Ras-protein equivalent to the second beta strand of the β-grasp fold and the polar residue at the C-terminus of the main helix. Hydrophilic residues are highlighted in cyan and the hydrophobic ones in yellow. Asterisks indicate the conserved positions of polar residues in the Ras-binding epitope. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
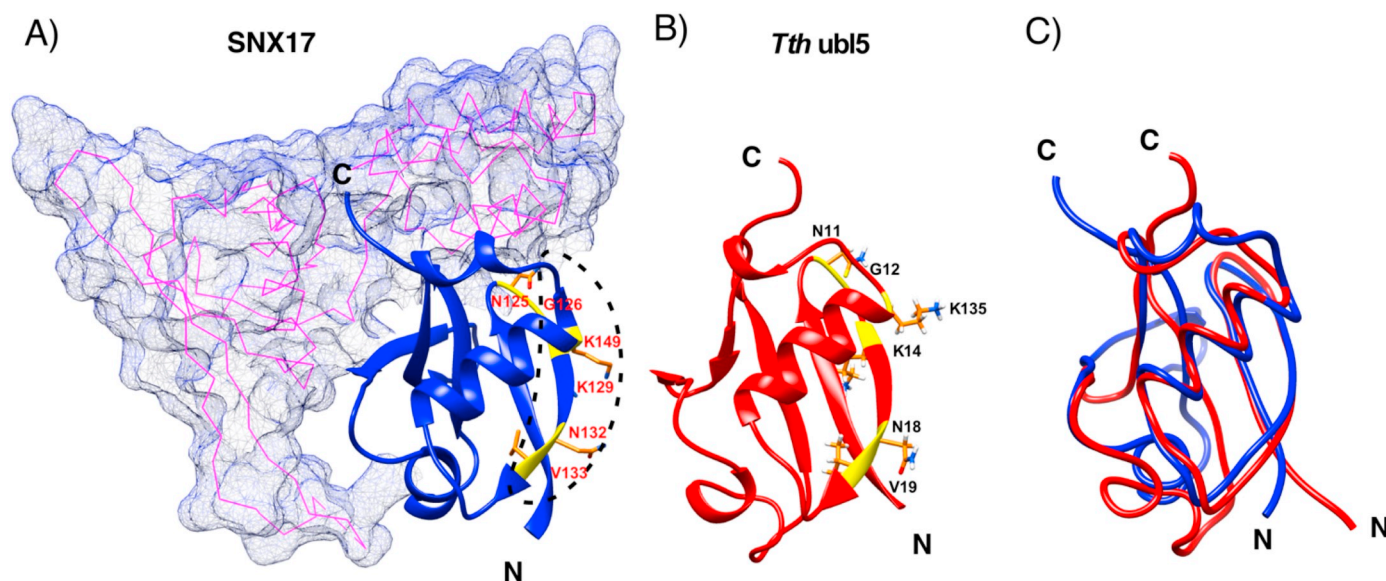
Fig. 6. Structural comparison of SNX17 (PDB ID: 4GXB) and *T.th*-ubl5. A) SNX17-UbL is shown in blue cartoons in relation with its host protein (lines-mesh). The Ras-binding epitope of SNX17 is indicated by the dotted line. B) *T.th*-ubl5 is depicted in red cartoons representation. Conserved residues between the two domains are highlighted in yellow. C) Structural superimposition of SNX17 and *T.th*-ubl5 (lowest energy conformer) performed with CHIMERA. N- and C-termini are indicated by letters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

membrane-localization modules found in several proteins associated with the cytoskeleton [42] [43]. FERMs are composed of three sub-domains 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 β_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,
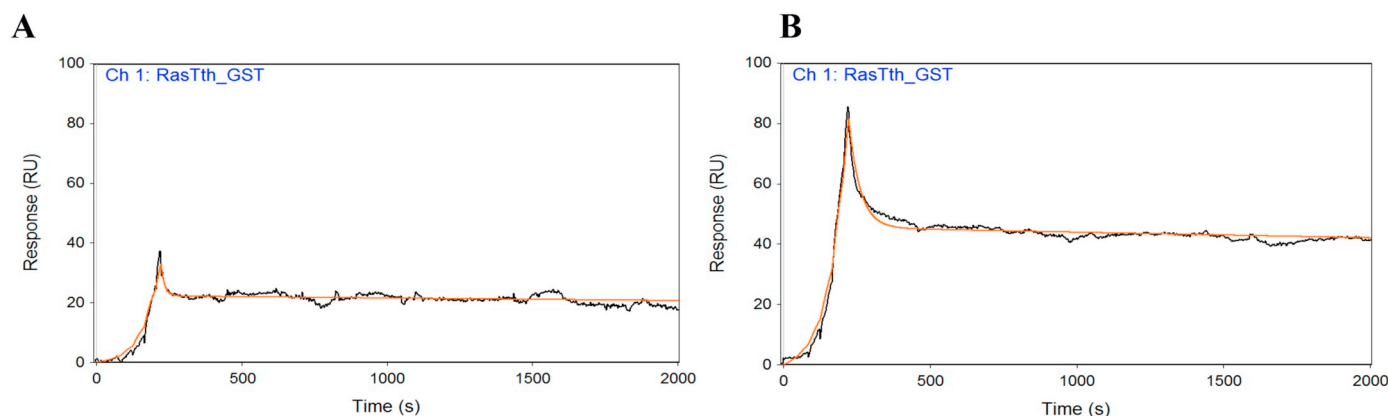

Fig. 7. SPR interaction experiment. GST-*T.th*-Ras in HSP buffer was injected on the sensor chip at a constant flow (30 μl/min). A FastStep procedure was used: the analyte was automatically diluted in HSP and injected by 6 serial doubling steps: 1) 0–40 s; 2) 41–80 s; 3) 81–120 s; 4) 121–160 s; 5) 161–200 s; 6) 201–220 s, where analyte concentrations were: 1) 0.0625 μM; 2) 0.125 μM; 3) 0.25 μM; 4) 0.5 μM; 5) 1 μM; 6) 2 μM (A), and 1) 0.25 μM; 2) 0.5 μM; 3) 1 μM; 4) 2 μM; 5) 4 μM; 6) 8 μM (B). The increase in RU relative to baseline indicates complex formation, whereas the decrease in RU after 220 s represents dissociation of GST-*T.th*-Ras from immobilized ligands after injection of buffer HSP. The sensorgrams were analysed using the SensiQ Qdat 4.0 program, using full fittings with 2 sites (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
ubq/ubl5 – UIM interactions. MD simulation results are summarized for each system (1 to 10). Number of LCF (Loss of Contacts per Frame) and FNB (Frames Not Bound) are reported, together with their average value over each run. Simulation replicas are indicated as _a and _b while "mut" denote a domain that was mutated in that particular system (see Fig. 3).

	Simulation	n° LCF	$\overline{LCF}$	% FNB	$\overline{\%FNB}$	Max contacts
1	*H.sap*-ubq/*H.sap*-UIM1_(1YX5)_a	0.61	0.82	1.73	2.12	28
	H.sap-ubq/*H.sap*-UIM1_(1YX5)_b	1.04		2.56		28
2	*H.sap*-ubq/*H.sap*-UIM1-mut_a	0.81	1.74	6.23	8.71	10
	H.sap-ubq/*H.sap*-UIM1-mut_b	2.67		11.19		17
3	*H.sap*-ubq/*T.th*-UIM1_a	0.23	0.46	0.71	2.06	18
	H.sap-ubq/*T.th*-UIM1_b	0.69		3.41		16
4	*T.th*-ubl5/*H.sap*-UIM1_a	3.69	2.15	22.42	13.15	10
	T.th-ubl5/*H.sap*-UIM1_b	0.61		3.88		11
5	*T.th*-ubl5/*T.th*-UIM1_a	4.36	7.51	10.39	13.74	22
	T.th-ubl5/*T.th*-UIM1_b	10.66		17.09		31
6	*T.th*-ubl5-mut/*T.th*-UIM1_a	1.62	1.42	5.61	6.05	14
	T.th-ubl5-mut/*T.th*-UIM1_b	1.23		6.49		15
7	*T.th*-ubl5/*H.sap*-UIM2_a	15.85	8.01	75.92	39.53	20
	T.th-ubl5/*H.sap*-UIM2_b	0.18		3.14		5
8	*T.th*-ubl5/*T.th*-UIM2_a	0.01	0.04	1.72	2.71	1
	T.th-ubl5/*T.th*-UIM2_b	0.07		3.70		2
9	*T.th*-ubl5-mut/*T.th*-UIM2_a	8.86	5.24	46.67	29.16	19
	T.th-ubl5-mut/*T.th*-UIM2_b	1.63		11.66		14
10	*H.sap*-ubq/*H.sap*-UIM2-mut_a	1.88	1.70	6.49	10.91	29
	H.sap-ubq/*H.sap*-UIM2-mut_a	1.53		15.34		10

the N-terminus region (res 1–34) showed high sequence similarity of 51.5%, further supporting our proposal that this domain contains the putative Ras-binding site.

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. 4HDO, 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* using H-Ras sequence as the query, found a single member of the small GTPase family (UNIPROT_I7M028) with good sequence conservation in the binding site (Q-D-T-Y-H) (Fig. S4). The rest of the matches were Rab proteins, which are not known to interact with UbL domains. In order to experimentally confirm the interaction between *T.th*-ubl5 and *T.th*-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-terminus, so as to investigate if its recruitment can occur before and/or concomitantly with the splicing processing. 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. 7) 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 K_D 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 K_D values (K_D = 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 BUBL1 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].

5. Conclusions

The biological role of the different domains of the BUBL locus and their combinations remains to be fully understood especially in relation with its exclusive conservation in early eukaryotes. Nonetheless, its inherent complexity and potential reactiveness, now together with the

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 ubl4 a conjugatable moiety and *T.th*-ubl5 the decoy recruiting Ras on to a larger ADP-ribosylation domain. Moreover, the strategical position of ubl4 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 parasitism 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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagen.2019.01.014.

References

[1] C.M. Pickart, Mechanisms underlying ubiquitination, Annu. Rev. Biochem. 70 (2001) 503–533, https://doi.org/10.1146/annurev.biochem.70.1.503.

[2] M. Hochstrasser, Origin and function of ubiquitin-like proteins, Nature 458 (2009) 422–429, https://doi.org/10.1038/nature07958.

[3] R.L. Welchman, C. Gordon, R.J. Mayer, Ubiquitin and ubiquitin-like proteins as multifunctional signals, Nat. Rev. Mol. Cell Biol. 6 (2005) 599–609, https://doi.org/10.1038/nrm1700.

[4] R. Hartmann-Petersen, C. Gordon, Integral UBL domain proteins: a family of proteasome interacting proteins, Semin. Cell Dev. Biol. 15 (2004) 247–259, https://doi.org/10.1016/j.semcdb.2003.12.006.

[5] L. Cappadocia, C.D. Lima, Ubiquitin-like protein conjugation: structures, chemistry, and mechanism, Chem. Rev. (2017), https://doi.org/10.1021/acs.chemrev.6b00737 acs.chemrev.6b00737.

[6] S. Jentsch, G. Pyrowolakis, Ubiquitin and its kin: how close are the family ties? Trends Cell Biol. 10 (2000) 335–342, https://doi.org/10.1016/S0962-8924(00)01785-2.

[7] A.M. Taherbhoy, B.A. Schulman, S.E. Kaiser, Ubiquitin-like modifiers, Essays Biochem. 52 (2012), http://essays.biochemistry.org/content/52/51.long (accessed July 13, 2017).

[8] M. Hu, P. Li, L. Song, P.D. Jeffrey, T.A. Chenova, K.D. Wilkinson, R.E. Cohen, Y. Shi, Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14, EMBO J. 24 (2005) 3747–3756, https://doi.org/10.1038/sj.emboj.

[9] J.F. Watkins, P. Sung, L. Prakash, S. Prakash2t, The saccharomyces cerevisiae DNA repair gene RAD23 encodes a nuclear protein containing a ubiquitin-like domain required for biological function, Mol. Cell. Biol. 13 (1993) 7757–7765 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC364847/pdf/molcellb00024-0569.pdf (accessed March 8, 2017).

[10] S.C. Upadhya, A.N. Hegde, A potential proteasome-interacting motif within the ubiquitin-like domain of parkin and other proteins, Trends Biochem. Sci. 28 (2003) 280–283, https://doi.org/10.1016/S0968-0004(03)00092-6.

[11] E. Sakata, S. Bohn, O. Mihalache, P. Kiss, F. Beck, I. Nagy, S. Nickell, K. Tanaka, Y. Saeki, F. Förster, W. Baumeister, Localization of the proteasomal ubiquitin receptors Rpn10 and Rpn13 by electron cryomicroscopy, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 1479–1484, https://doi.org/10.1073/pnas.1119394109.

[12] J.D. Aguirre, K.M. Dunkerley, P. Mercier, G.S. Shaw, Structure of phosphorylated UBL domain and insights into PINK1-orchestrated parkin activation, Proc. Natl. Acad. Sci. 114 (2017) 298–303, https://doi.org/10.1073/pnas.1613040114.

[13] D. Finley, B. Bartel, A. Varshavsky, The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis, Nature 338 (1989) 394–401, https://doi.org/10.1038/338394a0.

[14] V. Sauvé, A. Lilov, M. Seirafi, M. Vranas, S. Rasool, G. Kozlov, T. Sprules, J. Wang, J.-F. Trempe, K. Gehring, A Ubl/ubiquitin switch in the activation of Parkin, EMBO J. 34 (2015) 2492–2505, https://doi.org/10.15252/embj.201592237.

[15] A. Kumar, J.D. Aguirre, T.E.C. Condos, R.J. Martinez-Torres, V.K. Chaugule, R. Toth, R. Sundaramoorthy, P. Mercier, A. Knebel, D.E. Spratt, K.R. Barber, G.S. Shaw, H. Walden, Disruption of the autoinhibited state primes the E3 ligase parkin for activation and catalysis, EMBO J. 34 (2015) 2506–2521, https://doi.org/10.15252/embj.201592337.

[16] J.M. Winget, T. Mayor, The Diversity of Ubiquitin Recognition: Hot spots and Varied Specificity, Mol. Cell 38 (2010) 627–635, https://doi.org/10.1016/j.molcel.2010.05.003.

[17] B. Dassa, I. Yanai, S. Pietrokovski, New type of polyubiquitin-like genes with intein-like autoprocessing domains, Trends Genet. 20 (2004) 538–542, https://doi.org/10.1016/j.tig.2004.08.010.

[18] G. Amitai, O. Belenkiy, B. Dassa, A. Shainskaya, S. Pietrokovski, Distribution and function of new bacterial intein-like protein domains, Mol. Microbiol. 47 (2003) 61–73 http://www.ncbi.nlm.nih.gov/pubmed/12492854 (accessed February 15, 2017).

[19] M.W. Southworth, J. Yin, F.B. Perler, Rescue of protein splicing activity from a Magnetospirillum magnetotacticum intein-like element, Biochem. Soc. Trans. 32 (2004) 250–254 (doi:10.1042/).

[20] N. Oganesyan, I. Ankoudinova, S.-H. Kim, R. Kim, Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization, Protein Expr. Purif. 52 (2007) 280–285, https://doi.org/10.1016/j.pep.2006.09.015.

[21] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins Struct. Funct. Bioinforma. 59 (2005) 687–696, https://doi.org/10.1002/prot.20449.

[22] P. Güntert, Automated NMR structure calculation with CYANA, in: Protein NMR Tech., Humana Press, New Jersey, n.d.: pp. 353–378. doi:https://doi.org/10.1385/1-59259-809-9:353.

[23] P. Güntert, C. Mumenthaler, K. Wüthrich, Torsion angle dynamics for NMR structure calculation with the new program Dyana, J. Mol. Biol. 273 (1997) 283–298, https://doi.org/10.1006/jmbi.1997.1284.

[24] Y. Shen, A. Bax, Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks, J. Biomol. NMR 56 (2013) 227–241, https://doi.org/10.1007/s10858-013-9741-y.

[25] G. Cornilescu, F. Delaglio, A. Bax, Protein backbone angle restraints from searching a database for chemical shift and sequence homology, J. Biomol. NMR 13 (1999) 289–302 http://www.ncbi.nlm.nih.gov/pubmed/10212987 (accessed February 15, 2017).

[26] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules, Comput. Phys. Commun. 91 (1995) 1–41, https://doi.org/10.1016/0010-4655(95)00041-D.

[27] A. Bhattacharya, R. Tejero, G.T. Montelione, Evaluating protein structures determined by structural genomics consortia, Proteins Struct. Funct. Bioinforma. 66 (2006) 778–795, https://doi.org/10.1002/prot.21165.

[28] N.E. Hafsa, D. Arndt, D.S. Wishart, CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts, Nucleic Acids Res. 43 (2015) W370–W377, https://doi.org/10.1093/nar/gkv494.

[29] J. Yang, R. Yan, A. Roy, D. Xu, P.J.Y. Zhang, The I-TASSER suite: protein structure and function prediction, Nat. Methods 12 (2015) 7–8, https://doi.org/10.1038/nmeth.3213.

[30] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935, https://doi.org/10.1063/1.445869.

[31] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593, https://doi.org/10.1063/1.470117.

[32] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56, https://doi.org/10.1016/0010-4655(95)00042-E.

[33] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A.D. Mackerell Jr., CHARMM general force field:

a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, J. Comput. Chem. 31 (2010) 671–690, https://doi.org/10.1002/jcc.21367.

[34] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–8, 27–8 http://www.ncbi.nlm.nih.gov/pubmed/8744570 , Accessed date: 22 May 2017.

[35] H. Cheng, R.D. Schaeffer, Y. Liao, L.N. Kinch, J. Pei, S. Shi, B.-H. Kim, N.V. Grishin, ECOD: an evolutionary classification of protein domains, PLoS Comput. Biol. 10 (2014) e1003926, , https://doi.org/10.1371/journal.pcbi.1003926.

[36] J.S. Harrison, T.M. Jacobs, K. Houlihan, K. Van Doorslaer, B. Kuhlman, UbSRD: the ubiquitin structural relational database, J. Mol. Biol. 428 (2016) 679–687, https://doi.org/10.1016/j.jmb.2015.09.011.

[37] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780, https://doi.org/10.1093/molbev/mst010.

[38] O. O'Sullivan, K. Suhre, C. Abergel, D.G. Higgins, C. Notredame, 3DCoffee: combining protein sequences and structures within multiple sequence alignments, J. Mol. Biol. 340 (2004) 385–395, https://doi.org/10.1016/j.jmb.2004.04.058.

[39] L. Holm, P. Rosenstrom, Dali server: conservation mapping in 3D, Nucleic Acids Res. 38 (2010) W545–W549, https://doi.org/10.1093/nar/gkq366.

[40] I. Genovese, A. Fiorillo, A. Ilari, S. Masciarelli, F. Fazi, G. Colotti, Binding of doxorubicin to Sorcin impairs cell death and increases drug resistance in cancer cells, Cell Death Dis. 8 (2017) e2950, , https://doi.org/10.1038/cddis.2017.342.

[41] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132 http://www.ncbi.nlm.nih.gov/pubmed/7108955 (accessed August 8, 2017).

[42] A.H. Chishti, A.C. Kim, S.M. Marfatia, M. Lutchman, M. Hanspal, H. Jindal, S.C. Liu, P.S. Low, G.A. Rouleau, N. Mohandas, J.A. Chasis, J.G. Conboy, P. Gascard, Y. Takakuwa, S.C. Huang, E.J. Benz, A. Bretscher, R.G. Fehon, J.F. Gusella, V. Ramesh, F. Solomon, V.T. Marchesi, S. Tsukita, S. Tsukita, K.B. Hoover, D. Louvard, N.K. Tonks, J.M. Anderson, A.S. Fanning, P.J. Bryant, D.F. Woods, K.B. Hoover, The FERM domain: a unique module involved in the linkage of cytoplasmic proteins to the membrane, Trends Biochem. Sci. 23 (1998) 281–282, https://doi.org/10.1016/S0968-0004(98)01237-7.

[43] M.C. Frame, H. Patel, B. Serrels, D. Lietha, M.J. Eck, The FERM domain: organizing the structure and function of FAK, Nat. Rev. Mol. Cell Biol. 11 (2010) 802–814, https://doi.org/10.1038/nrm2996.

[44] J. Wojcik, J.-A. Girault, G. Labesse, J. Chomilier, J.-P. Mornon, I. Callebaut, Sequence analysis identifies a ras-associating (RA)-like domain in the N-termini of band 4.1/JEF domains and in the Grb7/10/14 adapter family, Biochem. Biophys. Res. Commun. 259 (1999) 113–120, https://doi.org/10.1006/bbrc.1999.0727.

[45] R. Ghai, M. Mobli, S.J. Norwood, A. Bugarcic, R.D. Teasdale, G.F. King, B.M. Collins, Phox homology band 4.1/ezrin/radixin/moesin-like proteins function as molecular scaffolds that interact with cargo receptors and Ras GTPases, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 7763–7768, https://doi.org/10.1073/pnas.1017110108.

[46] A.R. Gingras, W. Puzon-McLaughlin, M.H. Ginsberg, The structure of the ternary complex of Krev Interaction Trapped 1 (KRIT1) bound to both the Rap1 GTPase and the heart of glass (HEG1) cytoplasmic tail, J. Biol. Chem. 288 (2013) 23639–23649, https://doi.org/10.1074/jbc.M113.462911.

[47] R. Ghai, B.M. Collins, PX-FERM proteins: a link between endosomal trafficking and signaling? Small GTPases 2 (2011) 259–263, https://doi.org/10.4161/sgtp.2.5.17276.

[48] D.G. Woodside, D.K. Wooten, B.W. McIntyre, Adenosine diphosphate (ADP)-ribosylation of the guanosine triphosphatase (GTPase) rho in resting peripheral blood human T lymphocytes results in pseudopodial extension and the inhibition of T cell activation, J. Exp. Med. 188 (1998) 1211–1221 http://www.ncbi.nlm.nih.gov/pubmed/9763600 (accessed May 9, 2017).

[49] T.S. Vincent, J.E. Fraylick, E.M. McGuffie, J.C. Olson, ADP-ribosylation of oncogenic Ras proteins by Pseudomonas aeruginosa exoenzyme S in vivo, Mol. Microbiol. 32 (1999) 1054–1064, https://doi.org/10.1046/j.1365-2958.1999.01420.x.

[50] J. Qiu, M.J. Sheedlo, K. Yu, Y. Tan, E.S. Nakayasu, C. Das, X. Liu, Z.-Q. Luo, Ubiquitination independent of E1 and E2 enzymes by bacterial effectors, Nature 0 (2016) 1–17, https://doi.org/10.1038/nature17657.

[51] H. Kikuhara, M. Ogawa, H. Miyamoto, Y. Nikaido, S. Yoshida, Intracellular multiplication of Legionella pneumophila in Tetrahymena thermophila, J. UOEH 16 (1994) 263–275 http://www.ncbi.nlm.nih.gov/pubmed/7824817 (accessed May 9, 2017).

[52] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera?A visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612, https://doi.org/10.1002/jcc.20084.