Pak. j. life soc. Sci. (2017), 15(1): 51-59 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences
www.pjlss.edu.pk

RESEARCH ARTICLE

Characterization and *In-Silico* Studies on Ubiquitin Protein from Seeds of *Sisymbrium irio*

Binish Khaliq[1], Sadaf Iqbal[2], Sven Falke[3], Friedrich Buck[4], Aisha Munawar[5], Seema Mahmood[1], Christian Betzel[3] and Ahmed Akrem[1*]

[1]Botany Division, Institute of Pure and Applied Biology, Bahauddin Zakariya University, Multan, Pakistan

[2]Department of Chemistry, University of Karachi, Karachi, Pakistan

[3]University of Hamburg, Laboratory for Structural Biology of Infection and Inflammation, c/o Desy, Building 22a, Notkestrasse 85, Hamburg, Germany

[4]University Hospital Hamburg-Eppendorf, Institute for Clinical Chemistry, Martinistr 52, Hamburg, Germany

[5]Department of Chemistry, University of Engineering & Technology, G.T. Road, Lahore, Pakistan

ARTICLE INFO

Received: Dec 02, 2106
Accepted: Apr 23, 2017

Keywords
Circular dichroism
Dynamic light scattering
Homology modeling
Isoelectric focusing
Sisymbrium irio
Ubiquitin

***Corresponding Author:**
ahmedakrem@bzu.edu.pk

ABSTRACT

Ubiquitin protein was identified and characterized from the seeds of *Sisymbrium* (*S.*) *irio*. Results of LC-MS/MS spectrometry revealed that *S. irio* ubiquitin (*Si*u) shares around 77% sequence identity with already reported ubiquitin proteins from *Triticum aestivum*, *Avena sativa*, *Arabidopsis thaliana* and others. *Si*u revealed a molecular weight of 17 kDa on non-reduced SDS-PAGE while a reduced band of approximately 9 kDa in the presence of β-mercaptoethanol. A pI of 9.2 was calculated by isoelectric focusing measurement. Circular Dichroism (CD) spectroscopic studies showed that the secondary structure of *Si*u comprised of about 6.5% α-helix, 39.3% β-sheet, 0.7% turn, 53.5% random coil which indicates that it is mostly β-sheeted. Similarly, a hydrodynamic radius (R_H) of 2.44 nm was calculated by Dynamic Light Scattering (DLS) for monodispersive *Si*u which further infers the presence of a dimeric status of 9 kDa native protein in solution. Multiple sequence alignment of 20 different ubiquitin sequences revealed that ubiquitin sequence is well conserved in nature with very few amino acid differences, e.g. R42K, Q49D, and Q62D. Homology modeling of *Si*u (11-28, and 42-76) residues suggested overall conserved secondary and tertiary conformation and other structural motifs. This report describes the purification, characterization and structural analysis of an ubiquitin protein from seeds of *S. irio*.

INTRODUCTION

Ubiquitin is a globular and small protein present in all eukaryotic cells. The amino acid sequence of this protein is extremely well conserved from protozoan to vertebrates (Hochstrasser, 2000). Ubiquitin comprises of 76 amino acid residues and has a molecular mass of about 8.5 kDa (Thrower et al., 2000). The ubiquitinated proteins are well known for their degradation by 26S proteasome, one case being that more than four Ubs consists of a multi-Ub chain via lysine (K) 48 residues (Pickart, 2004). Genetic and molecular data in plants suggested that the Ub/26S pathway plays an important role in the response to stress. In all eukaryotic cells, Ubiquitin functions to modify proteins through a post-translational regulatory mechanism (Haglund and Dikic, 2005). The last four C-terminal amino acids (LRGG) extending from the compact structure, are responsible for its function (Mukhopadhyay and Riezman, 2007). Ubiquitin in the form of a signal molecule is covalently attached to a substrate protein which leads to many different outcomes. This attachment of ubiquitin to target proteins is responsible for the development of an isopeptide linkage between the C-terminal glycine and the epsilon amino group of lysine in the target proteins (Peng et al., 2003).

Additionally, this covalent conjugation of substrate proteins with ubiquitin (Ub) has developed as a powerful mechanism that controls the stability, function or localization of multiple proteins (Pickart, 2001). Similarly, other types of ubiquitination such as non-canonical ubiquitination and mono ubiquitination, are applied in many other cellular functions, including DNA damage repair, endocytosis, signal transduction and endosomal sorting (Jurado et al., 2008). Other than these important functions, ubiquitins are also involved in antigen process, apoptosis, immune response and inflammation ribosome biogenesis (Kurosu et al., 2007). Ubiquitination of target proteins needs the stepwise action of further three enzymes e.g., Ub-activating enzyme (E1), Ub-conjugating enzyme (E2) and Ub-ligase (E3) (Love et al., 2005).

Brassicaceae (*Cruciferae*) is an economically important plant family with wide spectrum agronomic traits (Aguilar and Wendland, 2003). London rocket *(Sisymbrium irio)* locally known as khub kalana in Pakistan is an annual herb and belongs to the family *Brassicaceae* (*Cruciferae*). It is grown in different parts of the Indo-Pak subcontinent and in the Middle East. *S. irio* is a minor oil crop; widely used as culinary and for medicines as remedies for different diseases e.g., inflammation, rheumatoid antipyretic, analgesic and antioxidant (Vohora et al., 1980). In the present study, *S. irio* ubiquitin (*Si*u) was analyzed thoroughly and a comprehensive molecular characterization is presented here.

MATERIALS AND METHODS

Plant material

Seeds of *S. irio* (Voucher number: BZBOT00110734) were taken from the Botanical garden, Bahauddin Zakariya University Multan, Pakistan.

Extraction and purification of ubiquitin

Seeds (5 g) of *S. irio* were finely grinded by pestle and mortar to a powder form and was homogenized in 50 ml citrate buffer (100 mM, pH 3.0) containing 1 mM of phenylmethylsulfonyl floride (PMSF) as protease inhibitor. The mixture was stirred continuously for five hours at 25°C. After stirring, the sample was centrifuged (Ogawa 6470, Japan) at 17,000 × g for 45 minutes and the pellet was discarded. The supernatant was filtered through Whatman filter paper (pore size 8 µm; EW-06648-46) to remove any particulate matter. The clear crude extract (50 ml) was concentrated to 10 ml using a concentrator with a 3 kDa cut off membrane (AMICON) and was applied onto a Hi-Load 16/60 Superdex 200 pg column using citrate buffer containing 150 mM NaCl. The protein was eluted with the same citrate buffer at a flow rate of 1.0 ml/min. Absorbance of the eluent was recorded at 280 nm. The fractions with maximum protein contents were combined and analyzed by SDS-PAGE. 0.01% NaN_3 was added to avoid any microbial growth. All the protein quantifications for the crude extract or the purified ubiquitin were calculated by applying a Nanodrop 2000c (Thermo Scientific, peqLab, Germany). 2 µl of protein solution were applied to the sensor while extraction buffer was used as blank. Absorbance was converted to corresponding quantifications by the help of Lambert and Beer equation as given below:

$A280 = \varepsilon * b * c$

Gel electrophoresis (SDS-PAGE)

The common procedure (Laemmli, 1970) was adopted for the preparation of the one dimensional 15% gels (E-VS10-SYS, omniPAGE mini-System, Germany) to visualize the protein samples. The gel was stained with either coomassie brilliant blue R-250 dye (Sigma Aldrich, CBBR-250) for visualizing the protein bands and molecular weight determination by using protein marker (Thermo Scientific; Catalogue # 22610).

Mass spectrometric analysis and protein identification

Gel bands stained with colloidal Coomassie dye were cut out and reduced with DTT (10 mM, 56°C, 30 min.). The protein was digested with trypsin (5 ng/µl; sequencing grade trypsin, Promega, USA) in 50 mM NH_4HCO_3 at 37°C for 16h. After digestion, the gel pieces were extracted in 50% acetonitrile/5% formic acid solution, the combined extracts were lyophilized in a vacuum concentrator and re-dissolved in 20 µl of 0.1% formic acid.

LC–MS/MS data were recorded by injecting the protein samples on a nano liquid chromatographic system (Dionex UltiMate 3000) coupled via electrospray-ionization (ESI) to an orbitrap mass spectrometer (Orbitrap Fusion, Thermo Scientific, Germany). Mass spectrometric data was recorded in positive ion mode. LC–MS/MS measurement was carried out in data dependent acquisition mode (DDA). MS/MS spectra were measured in the ion trap of the instrument.

LC–MS raw data were processed with software Proteome Discoverer 2.0 (Thermo Scientific, Germany). For ubiquitin identification, MS/MS spectra were screened with Sequest HT against the *Arabidopsis* and the plant Uniprot data base (www.uniprot.org: downloaded November 10, 2015) and a contaminant database (298 entries). The searches were performed using the following parameters: precursor mass tolerancee 10 ppm, fragment mass tolerance 0.2 Da, two missed cleavages allowed, carbamidomethylation of cysteine residues as fixed modification, oxidation of methionine residues as a variable modification. Identifications were validated manually.

Isoelectric focusing (IEF)

IEF was performed on 17 cm long and 0.5 mm gel thickness strip (pH 3–10, Sigma). Purified ubiquitin (505 µg) was loaded on a horizontal gel maintained at 28°C in dehydration buffer containing (8 M urea, 2%

CHAPS, 50 mM dithiothreitol (DTT), 0.2% Bio-Lyte ampholytes and 0.001% bromophenol blue) overnight. The pI markers (Sigma, USA) ranging from 3 to 10 pH were used to estimate the isoelectric pH of the proteins. Isoelectric focusing was performed in an IEF focusing cell (Bio-Rad). After IEF, the proteins were stained by coomassie brilliant blue.

Circular dichroism (CD) analysis

CD experiment was performed to determine the secondary structure of ubiquitin on CD6 dichrograph (Jobin Yvon, Longjumeau, France). Purified ubiquitin (0.2 mg/ml) was prepared in 100 mM citrate buffer, pH 3.0. The CD spectrum of ubiquitin was recorded in the far-UV-range between 195-260 nm at 25°C in a 1 mm path length quartz cell. The percentage of secondary structure of ubiquitin was determined by using Spectra Manager™ software (Jasco).

Dynamic light scattering (DLS) measurements

Dynamic light scattering (DLS) measurement was used to determine the size distribution profile of particles in solution. When the laser light passes through the solution, light scatters from the solution corresponding to the diffusion coefficient of particles by equating through the Stokes-Einstein equation (Brown, 1993). Analysis of the autocorrelation function in terms of particle size distribution was computed which tells about a monodisperse sample which should give a single exponential decay. A 15 µl sample of *Si*u protein was pipette out in a cuvette and dynamic light scattering (DLS) signals of the purified protein was recorded using a SpectroLight 300 (Xtal Concept) over a suitable period of time.

In-Silico analysis of *Si*u

The obtained peptide fragments were submitted to Quick 2D (Alva et al., 2016) server for the prediction of secondary structural elements. Each individual peptide was identified by LC–MS/MS and their possible combinations were searched using BLAST (Altschul et al., 1997) in the UniProtKB/Swiss-Prot database for their recognition. For multiple sequence alignment of amino acid sequences that are identical with our provided arranged sequence, a search was also performed applying BLAST in the UniProtKB/Swiss-Prot database and PDB as well for homology modelling. Non-redundant sequences from 20 different species were subsequently aligned using ClustalW by keeping the default parameters fixed (Thompson et al., 1994). The crystal structure of an ubiquitin from single cell fungi *Saccharomyces cerevisiae* (*Sc*-Ub_2, PDB-ID: 4NNJ_D) (Schäfer et al., 2014) was found as the best template by providing 90% sequence identity, 56% similarity and an expect value of 1e-24. The aligned sequences of *Si*u and *Sc*-Ub_2 were subjected to model building in Swiss-Model server (Biasini et al., 2014, Arnold et al., 2006). Models were built based on the target-template alignment using ProMod3 (Benkert et al., 2011). Coordinates which were conserved between template and target are copied from the template to the model. Fragment library was used to remodel insertions and deletions and side chains were then rebuilt. In the end, the conformation of the resulting structure was regularized by using a force field. The quality of the model was assessed by drawing Ramachandran plot using Procheck (http://services.mbi.ucla.edu/Procheck/) (Laskowski et al., 2001). Figure 5A-B containing structural details were rendered using Chimera (Pettersen et al., 2004).

RESULTS

Protein purification, identification and amino acid sequencing

Ubiquitin protein, isolated from seeds of *S. irio*, was purified by subjecting to size-exclusion chromatography in 100 mM citrate and 150 mM NaCl buffer, pH 3.0 to obtain the purified protein fractions. Fig. 1A shows the presence of three protein bands appeared at pH 3.0 while two distinct peaks obtained in the chromatogram as displayed in Fig. 1B. Pure protein fractions showing high absorbance at 280 nm were stored at 4°C. One dimensional SDS-PAGE was performed under both reducing (with β-mercaptoethanol) and non-reducing conditions to investigate the molecular weight and number of protein bands in *S. irio* extract as demonstrated in Fig. 1C. The results of SDS-PAGE showed that the purified protein has a molecular weight of ~17 kDa under non-reducing conditions (lane L1) and about 9 kDa under reducing conditions (lane L2). These results suggest the possibility of the presence of a dimer of the purified protein in solution. The purified protein was digested with trypsin and resulting peptides were subjected to LC–MS/MS mass spectrometry. Five peptides i.e. KTITLEVE SS (10), DTIDNVKAK (9), LIFAGKDL (8), EDGRT LADYN (10) and IDKESTLHLVLRLR (12) were identified in the protein of interest from the *S. irio* seeds which on BLAST search confirmed the ubiquitin identity as exemplified in Table 1. IEF showed single band on the IEF gel revealing the basic pI of 9.2 as illustrated in Fig. 1D.

CD spectroscopy

Fresh samples were dialyzed at a low salt concentration buffer and secondary structure was determined by using CD spectroscopy. The shape of the averaged spectrum is typical for proteins containing α-helix and β-sheet structure (Manavalan and Johnson, 1983) possessing two distinct minima at 195 nm and 260 nm. Fig. 2A shows the CD spectrum resulting in a calculated content of about 6.5% α-helix, 39.3% β-sheet, 0.7% turn, 53.5% random coil. The root-mean-square (RMS) values between the fitted curves (Reed and Kinzel, 1984) and the ubiquitin data was found as 5.11%.

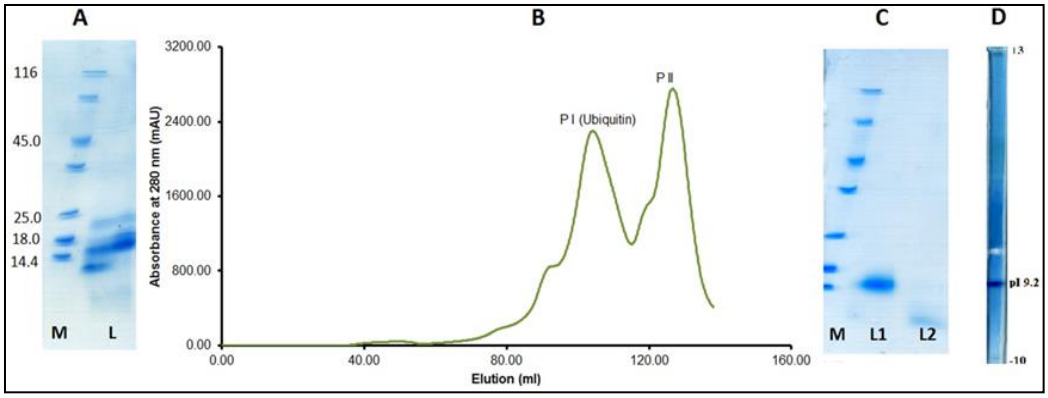

Fig. 1: Extraction, purification and isoelectric focusing of ubiqitin. (A) SDS-PAGE analysis of *S. irio* crude extract at pH 3.0 in lane L, (B) Size-exclusion chromatogram obtained from Superdex 200 column, (C) SDS-PAGE analysis of *S. irio* purified protein, comprising lane M for molecular weight standard, lane L1 for non-reducing condition and lane L2 for reducing condition. (D) Isoelectric focusing (IEF) of purified *S. irio* ubiquitin showing single band on strip with the alkaline pI of 9.2.

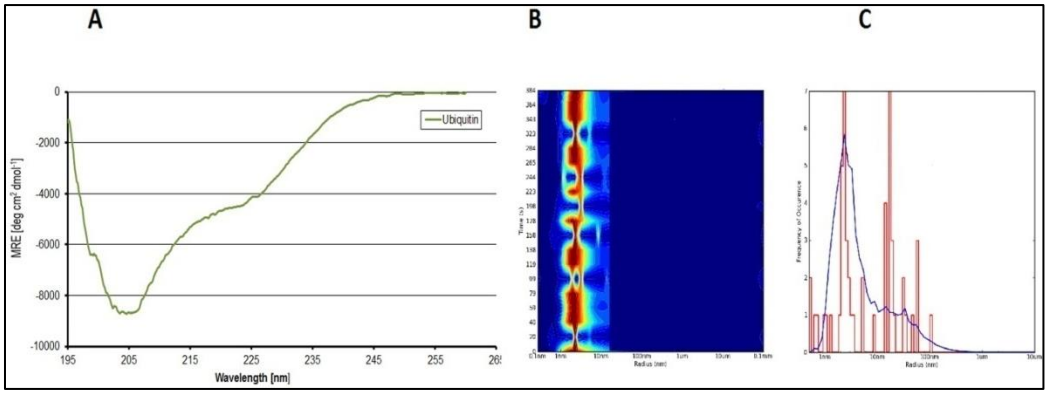

Fig. 2: Circular dichroism and dynamic light scattering of Ubiquitin. (A) Far-UV CD spectrum, (B) DLS and (C) the hydrodynamic radius (R_H) of ubiquitin purified from *S.irio* seeds.

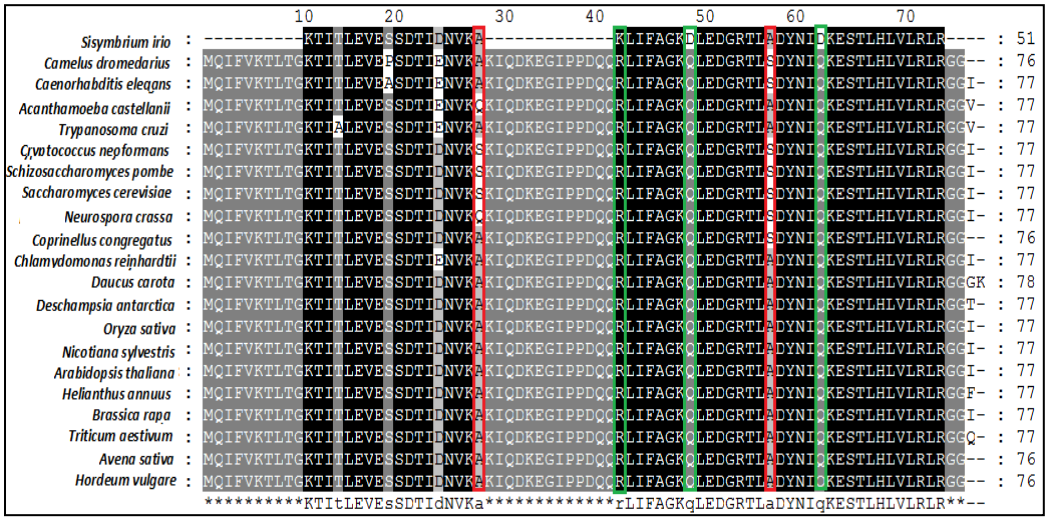

Fig. 3: Multiple sequence alignment of ubiquitin proteins from different species present in UniprotKB database. Black background showed overall conserved regions and grey background also revealed conserved regions but not obtained through LC–MS/MS in *Si*u sequence, - represents gaps whereas * indicates exactly identical residues. Red boxes show conserved changes while green boxes represent non-conserved changes.

Table 1: Sequence similarity of identified peptide fragments obtained from LC–MS/MS

Peptide fragments	% Identity	E-value	Protein name, organism & UniProtKB/Swiss-Prot ID: (Amino acid numbering)
KTITLEVESS	90	4×10^{-3}	polyubiquitin, *Arabidopsis thaliana*, Q39256.1 (89-98)
DTIDNVKAK	100	2.0×10^{-2}	polyubiquitin, *Arabidopsis thaliana*, Q39256.1 (99-107)
LIFAGKDL	88	8.4	polyubiquitin, *Caenorhabditis elegans*, P0CG71.1 (119-126)
EDGRTLADYN	100	7.0×10^{-4}	polyubiquitin, *Arabidopsis thaliana*, Q39256.1 (129-138)
IDKESTLHLVLRLR	93	3.0×10^{-6}	polyubiquitin, *Caenorhabditis elegans*, P0CG71.1 (137-150)
KTITLEVESS DTIDNVKAK LIFAGKDL EDGRTLADYN IDKESTLHLVLRLR	77	1.0×10^{-24}	Ubiquitin, *Triticum aestivum* or bread wheat, P69326.2 (11-74)

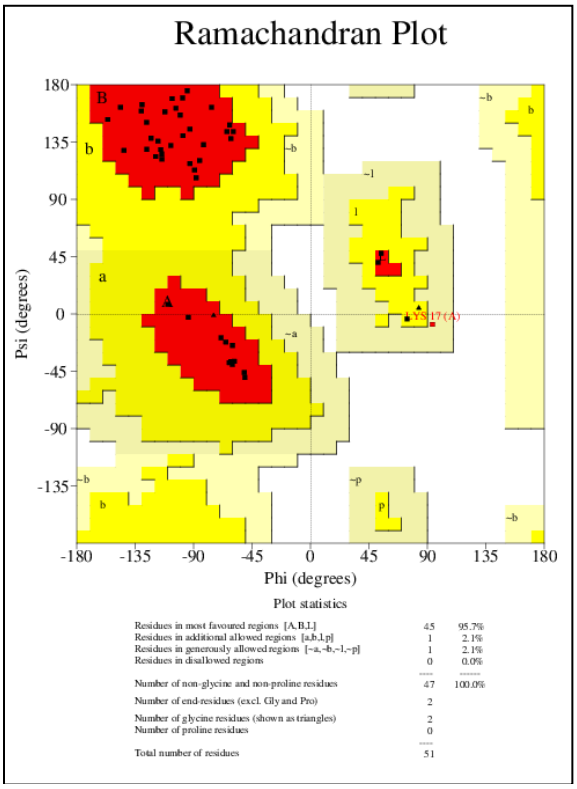

Fig. 4: Ramachandran plot of the modeled *Si*u structure indicating no amino acids in the disallowed region as an indication of good quality model.

DLS measurements

The *Si*u protein sample was subjected to DLS measurements after centrifugation, in order to remove unordered aggregates. Fig. 2B shows the DLS results of *Si*u protein revealing the monodispersive nature in solution, inferred from their single line signal with the hydrodynamic radius (R_H) of 2.44 (Fig. 2C). Although ubiquitins have also been reported as 9 kDa in native conformation as well, however, *Si*u appeared in dimeric form of 17 kDa under non-reduced conditions of SDS-PAGE. This R_H value of 2.44 provided further confirmation of the dimeric status of globular shape *Si*u in solution form.

In-Silico analysis of *Si*u

Fig. 3 demonstrates the multiple sequence alignment of *S. irio* ubiquitin (*Si*u) amino acid sequence with other homologues ubiquitin sequences present in the UniProtKB database. The analysis showed that the identified 51 amino acid sequence of *Si*u possess 77% identity with other ubiquitin sequences obtained from different species; *e.g.*, *Camelus dromedaries, Caenorhabditis elegans, Acanthamoeba castellanii, Trypanosoma cruzi, Cryptococcus neoformans, Schizosaccharomyces pombe, Saccharomyces cerevisiae, Neurospora crassa, Coprinellus congregates, Chlamydomonas reinhardtii, Daucus carota, Deschampsia antarctica, Oryza sativa, Nicotiana sylvestris, Arabidopsis thaliana, Helianthus annuus, Brassica rapa, Triticum aestivum, Avena sativa* and *Hordeum vulgare. Minor changes were observed in Si*u sequence at positions 42, 49 and 62. Structure and sequence of ubiquitin was recently reviewed in the literature which revealed that despite of few conservative changes; ubiquitin sequence is almost invariant from yeast to man (Aguilar et al., 2003). Presence of alanine residues in place of serine at positions 28 and 57 were considered as conservative changes because it was also observed in other sequences. *However, replacement of R42 by K42, Q49 and Q62 by D49 and D62, respectively are non-conserved changes and might be consider as the unique features of Si*u. The arranged amino acid sequence (1-51) of *Si*u was perfectly aligned with the amino acids (1-76) of *Saccharomyces cerevisiae* ubiquitin (*Sc*-Ub_2, PDB-ID: 4NNJ_D). The alignment induces two gaps in the *Si*u sequence corresponding to the amino acids # 1-10, and 29-41 of *Sc*-Ub_2.

A homology model of *Si*u with amino acids (11-28 and 42-76) was successfully generated by using the crystal structure information of *Sc*-Ub_2 (PDB-ID: 4NNJ_D). The quality of the model was checked by Procheck online tool which gave a Ramachandran plot for the modeled *Si*u as displayed in Fig. 4. Out of 51 modeled residues, Ramachandran plot showed 95.7% in the most favored region, 2.1% in allowed region, 2.1% in generously allowed and 0% in disallowed region.

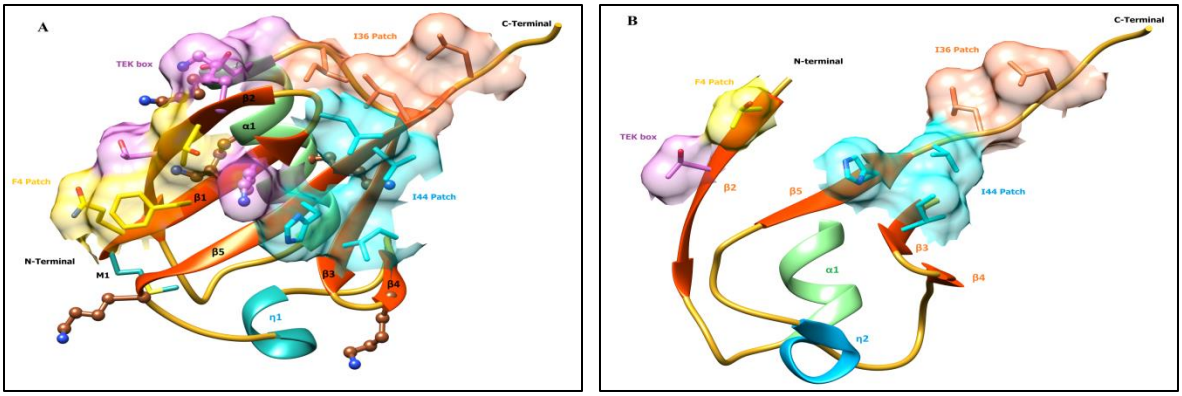

Fig. 5: Structural features of ubiquitin from *Sisymbrium irio*. (A) human erythrocytic ubiquitin (PDB-ID: 1UBQ) and (B) *S. irio* ubiquitin (*Si*u). Structural motifs include F4 patch yellow, TEK box orchid, I36 patch salmon, and I44 patch cyan in colors.

Therefore, the modeled *Si*u structure is good enough for further analysis. The crystal structure of human erythrocytic ubiquitin (heu) has been explained in detail and its structural features can be compared with the modeled *Si*u (Vijay-Kumar et al., 1987). In general, ubiquitin structure is compact in nature and only last six C-terminal residues (71-76) are unordered. Fig. 5A displays the crystal structure of heu showing all of its structural features. According to the human erythrocytic ubiquitin numbering (PDB-ID: 1UBQ), ubiquitin structure comprised of five β-strands holding residues 1-7, 11-17, 40-45, 48-50, and 65-72 (Fig. 5A, orange sheets). Furthermore, one α-helix comprising of residues 23-34 and one 3-10 helix consisting 56-59 residues (Fig. 5A, sea green and light blue helices). The most important feature of ubiquitin is its seven lysines (K6, K11, K27, K29, K33, K48 and K63) which cover all surfaces of ubiquitin and point into distinct directions (Fig. 5A, brown ball & sticks). K6 and K11 are located in the most important region and may lead towards conformational changes upon association with ubiquitin binding domains (UBDs) (Fig. 5A). As K27 is buried, linkage assembly through this residue would require localized changes in ubiquitin conformation (Geng et al., 2012). A hydrophobic surface comprising Q2, F4, and T12 (F4 patch, Fig. 5A, yellow surface) is obliged for cell division in yeast (*Saccharomyces cerevisiae*) (Sloper-Mould et al., 2001). This F4 patch interacts with the UBAN domain (Rahighi et al., 2009) and the ubiquitin-specific protease (USP) domain of deubiquitinating enzyme (DUBs) (Hu et al., 2002). In higher eukaryotes, a 3D motif that includes T12, T14, E34, K6, and K11 (TEK-box, Fig. 5A, orchid surface), is necessitate for mitotic degradation (Jin et al., 2008). A hydrophobic surface, centered on I36, L71 and L73 (I36 patch, Fig. 5A, salmon surface) can facilitates interactions between ubiquitin molecules. It is recognized by HECTE3s (Kamadurai et al., 2009), DUBs (Hu et al., 2002) and ubiquitin binding domains (UBDs) (Reyes-Turcu et al., 2006). Ubiquitin consists of hydrophobic surface, comprising of I44, L8, V70, and H68 (I44 patch, Fig. 5A, cyan surface) which is bound by the most UBDs, rendering it vital for cell division (Dikic et al., 2009; Sloper-Mould et al., 2001). Further, the β1/β2 loop containing L8 shows flexibility that is essential for recognition by ubiquitin-binding proteins (Lange et al., 2008). The modeled structure of *Si*u is explained in comparison with heu crystal structure as displayed in Fig. 5B. Due to incomplete sequence, the modeled *Si*u has missing β1 strand, lower part of α-helix (α1) and complete 3-10 helix (η1) (Fig. 5B).

DISCUSSION

Here we are describing the identification, purification and secondary conformation of an ubiquitin protein from seeds of *S. irio*. Isolated ubiquitin from *S. irio* has a molecular mass of around 17 kDa as visualized on SDS-PAGE (Fig. 1C, lane L1). Ubiquitins in plants exist either as monoubiquitins or polyubiquitins. Most of the ubiquitins have molecular masses in the range of 8.5 kDa and show similarities in their amino acid sequences. Isoelectric pH (pI) of isolated ubiquitin by isoelectric focusing experiment was found as 9.2 which are very near to the theoretical pI of reported ubiquitins. The amino acid sequence of *S. irio* ubiquitin was obtained by LC–MS/MS which provided five peptide fragments. Table 1 demonstrated the sequence similarity of identified peptide fragments obtained from LC–MS/MS analyzed by BLAST. Sequence similarity search in UniProtKB for individual peptides provided more than 88 % sequence identity with *Arabidopsis thaliana* polyubiquitin (Uniprot ID: Q39256.1) and *Caenorhabditis elegans* (Uniprot ID: P0CG71.1). However, when these five peptide fragments were

arranged according to the sequence numbering obtained from the previously identified ubiquitin proteins, it provides a combination of 57 amino acids sequence. The combined sequence showed 77% identity with *Triticum aestivum* (Uniprot ID: P69326.2) and *Avena sativa* (Uniprot ID: P69310.1).

CD spectroscopy data of *Si*u indicated the presence of both α-helix and β-sheet secondary structural elements in the native fold but mostly it comprised of β-sheet and random coils. The secondary structure estimation for the whole protein *Si*u by CD spectroscopy and the secondary structure prediction by Quick 2D for the obtained sequence of *Si*u matches well with each other. It also indicates secondary structure conservation within the reported ubiquitins with more content of β-sheet structure over α-helix. The obtained *Si*u sequence was submitted to Quick 2D for the prediction of secondary structural elements which showed β-sheet conformation for amino acid sequence ^{2}TITLEV^{7}, ^{20}LIF^{22} and ^{43}TLHLVLR^{49}. However, the peptide fragment ^{12}TIDNVKAK^{19} formed α-helix structure. Secondary structural elements of *Si*u revealed that β2 contains residues 12-17, α1 expresses 22-28 residues, β3 holds 43-46 amino acids, β4 comprises 48-49 residues, η2 encloses 56-60 amino acids, β5 constitutes 63-69 amino acids. Different structural motifs like F4 patch, TEK box, I36 patch and I44 patch are also available and conserved in *Si*u (Fig. 5B).

DLS showed hydrodynamic radius 2.44 with 17 kDa size of *Si*u protein was compared with the standard hydrodynamic radius of globular proteins (Erickson, 2009) which might infers the existence of an open conformation of either M1- or K63-linked diubiquitin dimer in solution. Structural characterization of different diubiquitin molecules revealed that different linkages between monomers result in different chain conformations; *i.e.,* compact and open. In the common model for K48-linked diubiquitin, the monomer chains interact through their I44 patches, and two such diubiquitins pack strongly in tetraubiquitin molecule (Tenno et al., 2004; Eddins et al., 2007). These types of association through K48-linked chains mostly adopt compact conformations. Similar to K48 linkages, K6- and K11-linked diubiquitin adopts compact conformations (Matsumoto et al., 2010; Bremm et al., 2010; Virdee et al., 2010) while M1- and K63-linked diubiquitin mostly exhibits open conformations (Komander et al., 2009; Weeks et al., 2009; Datta et al., 2009).

In order to analyze the conserved features of the ubiquitin sequence, a multiple sequence alignment of the combined peptide sequence of *S. irio* sequence and other homologous ubiquitins is performed (Fig. 2). Multiple sequence alignment suggested that ubiquitin proteins contain overall conserved sequences with very few non-conserved amino acids. *Si*u sequence is also found well conserved among different ubiquitins with conserved changes of A28 and A57 while non-conserved changes K42, D49, and D62. A homology model of the obtained sequence of *Si*u was constructed and compared with human ubiquitin for different structural features. Due to incomplete sequence, the modeled *Si*u has missing β1 strand, lower part of α-helix (α1) and complete 3-10 helix (η1) while other structural features are well conserved. The presence of seven lysines is the most unique feature of ubiquitins as already have been described and these lysines are very actively involved in a variety of their functions. In compliance to this, it is very interesting to note the mutation at position 42 where arginine has been replaced with another lysine, so altogether there has been the presence of eight lysines in the fragmented primary sequence of *Si*u. The presence of extra lysine has really made the things interesting to know its role in the interaction of *Si*u with other molecules. The presence of more lysine/s in gap areas of *Si*u primary sequence may also be possible. Similarly, the replacement of two glutamines with negatively charged aspartates is also point of consideration as it might result in conformational change of *Si*u and ultimately the corresponding function. These unique and typical changes really are tempting towards the detailed analysis of *Si*u molecular structure and also the interaction mechanisms with other molecules. These mutations may also be taken as an evidence of variation among very conserve ubiquitin sequence and quite possible that the discovery of more ubiquitins from different organisms lead towards the sequence variety and ultimate function diversity of these important proteins. Hence, the crystallization attempts of this protein are underway to solve the native three dimensional structure of this important protein.

Conclusions

A 17 kDa ubiquitin was identified, characterized and purified by LC–MS/MS and size-exclusion chromatography from the seeds of *S. irio*. It shares around 77% sequence identity with ubiquitin proteins from *Triticum aestivum*, *Avena sativa*, *Arabidopsis thaliana* and others. Secondary structure of *Si*u consisted of about 6.5% α-helix, 39.3% β-sheet, 0.7% turn, 53.5% random coil which indicates that it is rich in β-sheet. Result of Multiple sequence alignment of 20 different ubiquitin sequences showed that ubiquitin sequence is well conserved in nature with very few and unique amino acids mutations (*R42* by *K42*, *Q49* and *Q62* by *D49* and *D62*, respectively). Homology modeling of *Si*u (11-28 and 42-76) residues suggested the overall conserved secondary and tertiary structure and other structural motifs.

Authors' contribution

BK performed the lab work and collected the data, SI did the *In-Silico* studies, SF performed the CD

Spectroscopic studies, FB helped in identifying the protein through Mass Spectrometry, AM, SM helped in improving the language, CB provided the technical comments while AA conceive the idea and design the whole project as group leader.

Acknowledgements

We thank Higher Education Commission (HEC), Islamabad, Pakistan for the financial support as a part of Indigenous scholarship program and IRSIP program for the first author (HEC pin no. 112-22140-2MB1-565) to skill herself and to work in University of Hamburg, Germany as a part of her PhD program.

REFERENCES

Aguilar RC and B Wendland, 2003. Ubiquitin: not just for proteasomes anymore. Current Opinion in Cell Biology, 15: 184-190.

Altschul SF, TL Madden, AA Schäffer, J Zhang, Z Zhang, W Miller and DJ Lipman, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, 25: 3389-3402.

Alva V, SZ Nam, J Söding and AN Lupas, 2016. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. Nucleic Acids Research, 44: W410-W415.

Arnold K, L Bordoli, J Kopp and T Schwede, 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics, 22: 195-201.

Benkert P, M Biasini and T Schwede, 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics, 27: 343-350.

Biasini M, S Bienert, A Waterhouse, K Arnold, G Studer, T Schmidt, F Kiefer, TG Cassarino, M Bertoni and L Bordoli, 2014. Swiss-Model: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research, 42:W252-8.

Bremm A, SM Freund and D Komander, 2010. Lys11-linked ubiquitin chains adopt compact conformations and are preferentially hydrolyzed by the deubiquitinase Cezanne. Nature Structural and Molecular Biology, 17: 939-947.

Brown W, 1993. Dynamic light scattering: the method and some applications. Oxford University Press, USA.

Datta AB, GL Hura and C Wolberger, 2009. The structure and conformation of Lys63-linked tetraubiquitin. Journal of Molecular Biology, 392: 1117-1124.

Dikic I, S Wakatsuki and KJ Walters, 2009. Ubiquitin-binding domains - From structures to functions. Nature Reviews Molecular Cell Biology, 10: 659-671.

Eddins MJ, R Varadan, D Fushman, CM Pickart and C Wolberger, 2007. Crystal structure and solution NMR studies of Lys48-linked tetraubiquitin at neutral pH. Journal of Molecular Biology, 367: 204-211.

Erickson HP, 2009. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. Biological Procedures Online, 11: 32.

Geng F, S Wenzel and WP Tansey, 2012. Ubiquitin and proteasomes in transcription. Annual Review of Biochemistry, 81: 177-201.

Haglund K and I Dikic, 2005. Ubiquitylation and cell signaling. The EMBO Journal, 24: 3353-3359.

Hochstrasser M, 2000. Evolution and function of ubiquitin-like protein-conjugation systems. Nature Cell Biology, 2: E153-E157.

Hu M, P Li, M Li, W Li, T Yao, JW Wu, W Gu, RE Cohen and Y Shi, 2002. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. Cell, 111: 1041-1054.

Jin L, A Williamson, S Banerjee, I Philipp and M Rape, 2008. Mechanism of ubiquitin-chain formation by the human anaphase-promoting complex. Cell, 133: 653-665.

Jurado S, S Diaz-Trivino, Z Abraham, C Manzano, C Gutierrez and C D Pozo, 2008. SKP2A, an F-box protein that regulates cell division, is degraded via the ubiquitin pathway. The Plant Journal, 53: 828-841.

Kamadurai HB, J Souphron, DC Scott, DM Duda, DJ Miller, D Stringer, RC Piper and BA Schulman, 2009. Insights into Ubiquitin Transfer Cascades from a Structure of a UbcH5B~Ubiquitin-HECT NEDD4L Complex. Molecular Cell, 36: 1095-1102.

Komander D, F Reyes-Turcu, JD Licchesi, P Odenwaelder, KD Wilkinson and D Barford, 2009. Molecular discrimination of structurally equivalent Lys 63-linked and linear poly-ubiquitin chains. EMBO Reports, 10: 466-473.

Kurosu M, P Narayanasamy, K Biswas, R Dhiman and DC Crick, 2007. Discovery of 1, 4-didydroxy-2-naphthoate prenyltransferase inhibitors: new drug leads for multidrug-resistant gram-positive pathogens. Journal of Medicinal Chemistry, 50: 3973-3975.

Laemmli UK, 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227: 680-685.

Lange OF, NA Lakomek, C Farès, GF Schröder, KF Walter, S Becker, J Meiler, H Grubmüller, C Griesinger and BL DeGroot, 2008. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science, 320: 1471-1475.

Laskowski R, M MacArthur and J Thornton, 2001. PROCHECK: validation of protein structure coordinates. International Tables of Crystallography, Vol. F. Crystallography of Biological Macromolecules. Kluwer Academic Publishers, The Netherlands, 722-725.

Love CG, AJ Robinson, GA Lim, CJ Hopkins, J Batley, G Barker, GC Spangenberg and D Edwards, 2005. *Brassica* ASTRA: an integrated database for *Brassica* genomic research. Nucleic Acids Research, 33: D656-D659.

Manavalan P and WC Johnson, 1983. Sensitivity of circular dichroism to protein tertiary structure class. Nature, 305: 831–832.

Matsumoto ML, KE Wickliffe, KC Dong, C Yu, I Bosanac, D Bustos, L Phu, DS Kirkpatrick, SG Hymowitz and M Rape, 2010. K11-linked polyubiquitination in cell cycle control revealed by a K11 linkage-specific antibody. Molecular Cell, 39: 477-484.

Mukhopadhyay D and H Riezman, 2007. Proteasome-independent functions of ubiquitin in endocytosis and signaling. Science, 315: 201-205.

Peng J, D Schwartz, JE Elias, CC Thoreen, D Cheng, G Marsischky, J Roelofs, D Finley and SP Gygi, 2003. A proteomics approach to understanding protein ubiquitination. Nature Biotechnology, 21: 921-926.

Pettersen EF, TD Goddard, CC Huang, GS Couch, DM Greenblatt, EC Meng and TE Ferrin, 2004. UCSF Chimera-a visualization system for exploratory research and analysis. Journal of Computational Chemistry, 25: 1605-1612.

Pickart CM, 2001. Mechanisms underlying ubiquitination. Annual Review of Biochemistry, 70: 503-533.

Pickart CM, 2004. Back to the future with ubiquitin. Cell, 116: 181-190.

Rahighi S, F Ikeda, M Kawasaki, M Akutsu, N Suzuki, R Kato, T Kensche, T Uejima, S Bloor and D Komander, 2009. Specific recognition of linear ubiquitin chains by NEMO is important for NF-κB activation. Cell, 136: 1098-1109.

Reed J and V Kinzel, 1984. Near-and far-ultraviolet circular dichroism of the catalytic subunit of adenosine cyclic 5'-monophosphate dependent protein kinase. Biochemistry, 23: 1357-1362.

Reyes-Turcu FE, JR Horton, JE Mullally, A Heroux, X Cheng and KD Wilkinson, 2006. The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. Cell, 124: 1197-1208.

Schäfer A, M Kuhn and H Schindelin, 2014. Structure of the ubiquitin-activating enzyme loaded with two ubiquitin molecules. Acta Crystallographica Section D: Biological Crystallography, 70: 1311-1320.

Sloper-Mould KE, JC Jemc, CM Pickart and L Hicke, 2001. Distinct functional surface regions on ubiquitin. Journal of Biological Chemistry, 276: 30483-30489.

Tenno T, K Fujiwara, H Tochio, K Iwai, EH Morita, H Hayashi, S Murata, H Hiroaki, M Sato and K Tanaka, 2004. Structural basis for distinct roles of Lys63-and Lys48-linked polyubiquitin chains. Genes to Cells, 9: 865-875.

Thompson JD, DG Higgins and TJ Gibson, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22: 4673-4680.

Thrower JS, L Hoffman, M Rechsteiner and CM Pickart, 2000. Recognition of the polyubiquitin proteolytic signal. The EMBO Journal, 19: 94-102.

Vijay-Kumar S, CE Bugg and WJ Cook, 1987. Structure of ubiquitin refined at 1.8 Åresolution. Journal of Molecular Biology, 194: 531-544.

Virdee S, Y Ye, DP Nguyen, D Komander and JW Chin 2010. Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. Nature Chemical Biology, 6: 750-757.

Vohora S, S Naqvi and I Kumar, 1980. Antipyretic, analgesic and antimicrobial studies on *Sisymbrium irio*. Planta Medica, 38: 255-259.

Weeks SD, KC Grasty, L Hernandez-Cuebas and PJ Loll, 2009. Crystal structures of Lys-63-linked tri-and di-ubiquitin reveal a highly extended chain architecture. Proteins: Structure, Function, and Bioinformatics, 77: 753-759.