

Sequence and structure evolved separately in a ribosomal ubiquitin variant

André Catic^{1,5,7}, Zhen-Yu J Sun^{2,7},
Daniel M Ratner^{3,4}, Shahram Misaghi^{1,6},
Eric Spooner¹, John Samuelson⁴,
Gerhard Wagner^{2,*} and Hidde L Ploegh^{1,*}

¹Department of Biology, Whitehead Institute for Biomedical Research and Massachusetts Institute of Technology, Cambridge, MA, USA,

²Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA, ³Section of Infectious Diseases, Boston Medical Center, Boston, MA, USA and ⁴Department of Molecular and Cell Biology, Boston University Goldman School of Dental Medicine, Boston, MA, USA

Encoded by a multigene family, ubiquitin is expressed in the form of three precursor proteins, two of which are fusions to the ribosomal subunits S27a and L40. Ubiquitin assists in ribosome biogenesis and also functions as a post-translational modifier after its release from S27a or L40. However, several species do not conserve the ribosomal ubiquitin domains. We report here the solution structure of a distant variant of ubiquitin, found at the N-terminus of S27a in *Giardia lamblia*, referred to as GIUb_{S27a}. Despite the considerable evolutionary distance that separates ubiquitin from GIUb_{S27a}, the structure of GIUb_{S27a} is largely identical to that of ubiquitin. The variant domain remains attached to S27a and is part of the assembled holoribosome. Thus, conservation of tertiary structure suggests a role of this variant as a chaperone, while conservation of the primary structure—necessary for ubiquitin's function as a post-translational modifier—is no longer required. Based on these observations, we propose a model to explain the origin of the widespread ubiquitin superfold in eukaryotes. *The EMBO Journal* (2007) 26, 3474–3483. doi:10.1038/sj.emboj.7601772; Published online 28 June 2007
Subject Categories: proteins; structural biology
Keywords: evolution; ribosome; recombination; structure; ubiquitin

Introduction

Ubiquitin plays a central role in the biology of eukaryotes and strong sequence conservation is a hallmark of this post-

*Corresponding authors. G Wagner, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. Tel.: +1 617 726 6377; Fax: +1 617 724 2662; E-mail: wagner@hms.harvard.edu or HL Ploegh, Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA; Tel.: +1 617 324 1878; Fax: +1 617 452 3566; E-mail: ploegh@wi.mit.edu

⁵Present address: Harvard Stem Cell Institute, Massachusetts General Hospital, Boston, MA, USA

⁶Present address: Genentech Inc., San Francisco, CA, USA

⁷These authors contributed equally to this work

Received: 4 April 2007; accepted: 22 May 2007; published online: 28 June 2007

translational modifier (Nei and Rooney, 2005). In yeast, as well as in most higher eukaryotes, ubiquitin is expressed in the form of three different precursors. First, as a linear fusion protein consisting of five or more ubiquitin copies, linked in head-to-tail configuration (polyubiquitin). Second, ubiquitin is encoded as an N-terminal fusion to the ribosomal proteins S27a and L40, where it acts as a chaperone and assists in the formation of the holoribosome (Finley *et al*, 1989; Redman and Rechsteiner, 1989; Spence *et al*, 2000). The three ubiquitin moieties in these precursors are identical in amino-acid sequence within a species and they are also highly conserved between species. Regulated proteolysis by ubiquitin-specific proteases processes ubiquitin to its mature form, a 76-residue polypeptide terminating in a di-glycine motif. As a covalent modifier, ubiquitin is then involved in protein degradation, in intracellular transport, and in epigenetic regulation (Kerscher *et al*, 2006). The majority of mature ubiquitin in a cell derives from the ribosomal domains (based on data in budding yeast under standard growth conditions; Finley *et al*, 1989).

Most protists (e.g., *Giardia lamblia*, *Entamoeba histolytica*, *Trichomonas vaginalis*, *Trypanosoma cruzi*, *Plasmodium falciparum*, *Toxoplasma gondii*) and some metazoa (e.g., *Caenorhabditis*) do not harbor a conserved ubiquitin domain as a fusion with S27a, but instead encode a domain of similar length with unrelated amino-acid sequence (Jones and Candido, 1993; Catic and Ploegh, 2005). We and others have speculated how both selective pressure and recombination with the invariant polyubiquitin locus contribute to conservation of all three ubiquitin genes—a mechanism that appears to be leaky, especially for the S27a-linked ubiquitin domain (Catic and Ploegh, 2005; Nei and Rooney, 2005). Ribosomal ubiquitin presumably evolved by insertion of a DNA stretch from the invariant polyubiquitin gene in frame with S27a and L40. The presence of other ubiquitin fusions in some species (Archibald *et al*, 2003), together with the fact that S27a and L40 lack an additional N-terminal domain in archaea (which do not encode ubiquitin genes), support this model. Therefore, we assume that in eukaryotic species with a divergent domain preceding S27a, this moiety represents a deteriorated ubiquitin gene, a 'ubiquitin variant'.

Figure 1 depicts a multiple sequence alignment of naturally occurring fusion proteins with the N-terminal domains (referred to as Ub_{S27a}) at the top and the C-terminal ribosomal S27a polypeptides at the bottom. The depicted species express ubiquitin variants at the N-terminus, except for *Homo sapiens* and *Saccharomyces cerevisiae*, both of which encode a conserved bona fide ubiquitin (hereafter simply referred to as ubiquitin) at this position. The C-terminal ribosomal S27a polypeptides themselves display a high degree of conservation, compared to the N-terminal portions. In fact, some ubiquitin variants only share similar lengths, and show conservation of few amino acids when compared to ubiquitin. In particular, the sequences of ubiquitin variants of closely related species do not reflect evolutionary distance from their last common ancestor, which suggests that the

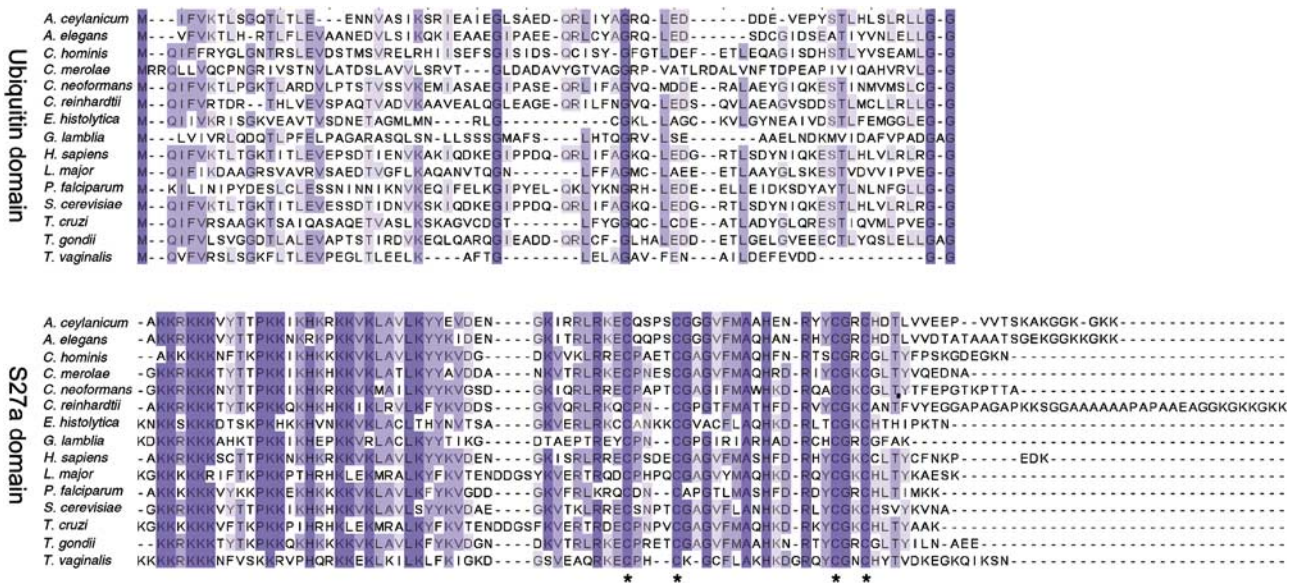


Figure 1 Multiple sequence alignment of ubiquitin-S27a fusion proteins. Shown are the ubiquitin domains Ub_{S27a} (top) and the ribosomal domains S27a (bottom) of 15 different species. *H. sapiens* and *S. cerevisiae* express conserved ubiquitin, all other species depicted encode ubiquitin variants. The asterisks indicate a conserved putative zinc-binding motif in the S27a domain (Chan *et al*, 1995).

primary structure of these proteins is not conserved (Supplementary Figures 1 and 2). Nonetheless, the isolated presence of invariant columns of glycines and hydrophobic side chains prompted us to speculate that the secondary and tertiary structures of these variants might be more similar than suggested by their sequences.

We were intrigued by the finding of non-conserved ubiquitin variants and by the fact that all ribosomal S27a or L40 polypeptides in eukaryotes contain N-terminal extensions of 60–80 amino acids, regardless of how different they are in sequence. We chose the binucleate, microaerophilic protist *G. lamblia*, a deeply divergent eukaryote, whose genome has been sequenced in its entirety, as a model to solve the structure and propose a function for its S27a-linked ubiquitin variant domain (referred to as GIUb_{S27a}).

Results

The variant GIUb_{S27a} has a typical ubiquitin fold

Human ubiquitin and GIUb_{S27a} share little sequence identity (18%) and no significant overall similarity according to ClustalX or BLAST2Seq, using either of the available substitution matrices (Thompson *et al*, 1997; Tatusova and Madden, 1999). However, as assessed by an ensemble of NMR structures (backbone atom r.m.s.d. 0.76 Å), their secondary structures align very closely and show the consensus sequence of ββ α ββ(α)β, typical for the ubiquitin-like fold (Figure 2A). In addition, a similar alternating pattern of polar and non-polar residues is apparent in both of these proteins. Such a ‘striped’ sequence is also evident in other ubiquitin fold proteins (Jentsch and Pyrowolakis, 2000; Lake *et al*, 2001; Buchberger, 2002; Supplementary Figure 3). To solve the structure of GIUb_{S27a}, we had to overcome solubility and aggregation problems by fusing the protein to hydrophilic terminal domains (see Materials and methods). The structure shown in Figure 2 is that of the GIUb_{S27a} protein

alone. The three-dimensional (3D) ribbon models of GIUb_{S27a} and ubiquitin show close similarity (Figure 2B and C), and both molecules have a comparable overall structure (Figure 2D). However, only four side chains on the surface have conserved charges. Even more striking, 59% of the amino acids in GIUb_{S27a} are hydrophobic and its surface is also remarkably non-polar (43.3% or 2327/5381 Å²). In contrast, 46% of the residues of ubiquitin—but only 28.7% (1390/4841 Å²) of its surface—are hydrophobic. Some of the most prominent differences between both sequences are present around the main α-helix. Four out of seven helical side chains facing the outer surface of ubiquitin are charged (E24, K29, D32, and K33), whereas no electrostatic potential exists at the corresponding side chains in GIUb_{S27a}. Furthermore, K27, the only conserved lysine in most other ubiquitin variants, is absent from GIUb_{S27a}. The ε-amino side chain of K27 in ubiquitin does not engage in formation of isopeptide-linked polyubiquitin chains. Rather, it appears that this charged residue is stabilizing the overall ubiquitin fold by forming a salt bridge with D52 (Vijay-Kumar *et al*, 1987). However, this interaction is dispensable for ubiquitin’s function (and likely for its structure, too) under permissive conditions, as shown by site-directed mutagenesis in budding yeast (Sloper-Mould *et al*, 2001). The assembly of the ubiquitin fold is nucleated by the highly conserved first β-strands and the α-helix. Rapid condensation into the final structure is driven by entropy—avoiding an intermediate state—and depends on the presence of a hydrophobic core (Went and Jackson, 2005). Three leucines in particular, L43, L50, and L56, are involved in wrapping the C-terminal β-strands 3 and 4 of ubiquitin around the initially formed α-helix. All three hydrophobic side chains are conserved in GIUb_{S27a} (Figure 2A).

Taken together, the NMR analysis reveals a close structural similarity between ubiquitin and GIUb_{S27a}, whereas the electrostatic and chemical properties of their surfaces differ

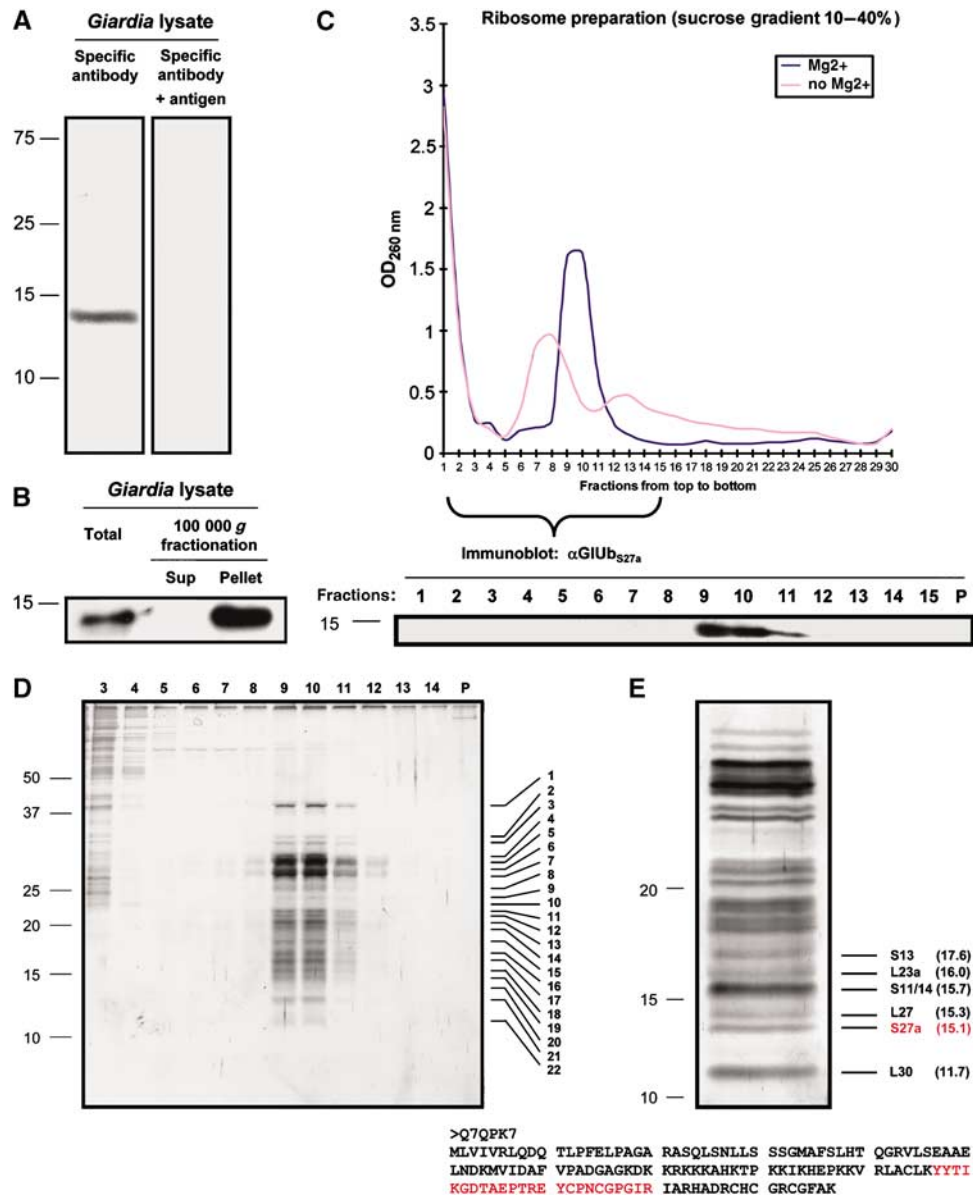


Figure 3 (A) GIUb_{S27a} appears as a single protein in *G. lamblia* lysate, in size consistent with a fusion to S27a. Whole-cell *G. lamblia* lysate (20 µg protein content per lane) was analyzed with polyclonal anti-GIUb_{S27a} antibodies by immunoblot. The serum reacts with a single protein (left lane), close to the predicted size of the GIUb_{S27a}-S27a fusion (15.1 kDa). No signal is detected at the predicted size of the GIUb_{S27a} domain alone (7 kDa). Incubation of the immunoblot with recombinantly expressed GIUb_{S27a} (10 µg) competes with detection of the protein (right lane). Pretreatment of trophozoites with the proteasome inhibitor ZL₃VS (50 µM, 2 h) did not alter the results (not shown). Immunoblots were resolved by reducing SDS-PAGE (12%). Molecular weight markers are indicated at the left (in kDa). (B) GIUb_{S27a} is enriched in the high-molecular-weight fraction. Anti-GIUb_{S27a} immunoblot of fractionated *G. lamblia* lysate. 30 µg of proteins were loaded per lane. GIUb_{S27a} is enriched in the sedimented fraction of *G. lamblia* lysate after ultracentrifugation (100 000 g for 1 h), but is absent from the supernatant. Samples were resolved by reducing SDS-PAGE (11%). (C) GIUb_{S27a} is associated with RNA-rich fractions, consistent with ribosome association. *G. lamblia* lysate was fractionated on 10–40% sucrose gradients, divided into equal fractions by volume and analyzed by UV spectroscopy (top panel). Addition of Mg²⁺ stabilizes the main RNA peak and we used such treated samples for the subsequent analyses. The top 15 fractions and the solubilized pellet (P) were assayed by anti-GIUb_{S27a} immunoblots (bottom panel), using equal volumes of each fraction (20 µl). The presence of GIUb_{S27a} coincides with the RNA peak in fractions 9, 10, and 11. Samples were resolved by reducing SDS-PAGE (11%). (D) Fractions 9–11 contain large and small ribosomal subunits, indicative of the assembled holoribosome. The previously obtained fractions (C) were resolved by reducing SDS-PAGE (11%) and silver stained. A total of 22 bands were excised, trypsinized and analyzed by MS/MS. All but one peptides identified are of ribosomal origin (see Table I). (E) Large-scale preparation of fraction 9 identifies S27a at the same molecular weight as GIUb_{S27a}. 70 µg protein lysate of fraction 9 was resolved by reducing SDS-PAGE (11%), silver stained, and individual polypeptides were analyzed by MS/MS. Excised bands and proteins identified are indicated at the right, with theoretical molecular weight of the full-length proteins in parentheses. We found peptides representing 34.3% of the C-terminal S27a domain (bottom panel, in red). The corresponding silver-stained protein (indicated with red) runs at the same relative position compared to the molecular weight markers as does the GIUb_{S27a} domain in the immunoblot shown in panel A. In addition, electrophoretic mobility in comparison with neighboring ribosomal proteins is consistent with a combined weight of 15.1 kDa of the GIUb_{S27a}-S27a fusion (SWISSPROT designation Q7QPK7).

proteolytically cleaved from S27a by virtue of the presence of a di-glycine-like motif at the junction between these two polypeptides. In ubiquitin and in ubiquitin-like molecules, this motif represents the P1 and P2 site for specific proteases that act on the precursors (Kerscher *et al*, 2006). Isolation of high-molecular-weight complexes by ultracentrifugation shows an enhancement of the anti-GlUb_{S27a} signal (Figure 3B), in agreement with it being fused to S27a and thus to the ribosome. To further demonstrate this association, we isolated ribosomes from *G. lamblia* by sucrose density gradient centrifugation. A comparison of the different fractions obtained by this method shows a perfect correlation between the anti-GlUb_{S27a} signal in immunoblots and the peak in RNA concentration, as determined by absorption at 260 nm (Figure 3C). Identification of the individual proteins within these fractions by silver staining and by tandem mass spectroscopy (MS/MS) confirmed that fractions 9 through 11 represent the assembled holoribosome, containing peptides of the small and the large subunit (Figure 3D and Table I). In fact, only one protein was found that has no apparent ribosomal homolog in higher eukaryotes (Q7QVJ9, theoretical pI 9.12, molecular weight 17.3 kDa). Since this protein has no sequence similarity to any previously defined domains, it might represent a *G. lamblia*-specific ribosomal peptide. Using a large-scale preparation of fraction 9, we also identified the S27a polypeptide with 34.3% sequence coverage (Figure 3E). Importantly, the band analyzed in the silver-stained SDS-PAGE has the same apparent molecular mass as the protein identified with the anti-GlUb_{S27a} serum (Figure 3A). Hence, with immunoblots for the N-terminal domain and mass spectroscopy for the C-terminal portion, we have confirmed that GlUb_{S27a} is fused to S27a and part of the fully assembled holoribosome, unlike Ub_{S27a} domains in yeast and in mammals (Finley *et al*, 1989; Redman and Rechsteiner, 1989; Louie *et al*, 1996; Vladimirov *et al*, 1996). We found no proof for a free, monomeric version of GlUb_{S27a}.

Table I Ribosomal proteins identified by MS/MS in the excised bands 1–22 (see Figure 3D)

Band	Ribosomal subunits
1	L3
2	L1
3	L5
4	S2, S3A, S4
5	S6
6	L7
7	L8
8	S5
9	L10A
10	L10, S3
11	L13, L19B, S11
12	L13A, S7
13	L6, L18, L18E, S5E
14	L11, L12, L22E, S8
15	L17, L21E
16	L35, S13, S13E, S16
17	L7A, L15, L23A, L27A, S15
18	L26, L32, S11/14, S17, S23
19	L14, L23, S10
20	L27, S15A, S24
21	L35A, S12, S20, S26
22	L7a, L24, L30, L37, L37, L37AE, L40, P2, S21E, S25, S27

No evidence for GlUb_{S27a}-specific proteases

G. lamblia encodes a single ubiquitin moiety instead of a polyubiquitin gene (Kreber *et al*, 1994), whereas both ribosomal domains GlUb_{S27a} and GlUb_{L40} harbor ubiquitin variants (Catic and Ploegh, 2005). Nevertheless, despite more than three billion years of combined divergence from their last common ancestor, budding yeast and *G. lamblia* possess remarkably similar ubiquitin conjugating enzymes, covering all biological aspects of ubiquitination (Supplementary Figure 4A and B). As expected, a recent study showed that ubiquitin is active as a post-translational modifier in *G. lamblia* (Gallego *et al*, 2007). In addition, this protist also encodes seven homologs to deubiquitinating enzymes (Supplementary Figure 4C). These are presumably necessary to remove the asparagine residue at the C-terminus of the ubiquitin precursor, in order to expose its di-glycine motif, and also to allow reversal of ubiquitination. In yeast as well as in higher eukaryotes, these proteases are furthermore required to remove the ubiquitin domains from the ribosomal S27a or L40 polypeptides, after they have completed their assistance in ribosome assembly. To rule out involvement of the seven putative deubiquitinating proteases in processing of GlUb_{S27a}, we generated an electrophilic probe (Borodovsky *et al*, 2002) of this ubiquitin variant. A covalent inhibitor based on ubiquitin reacts with several proteins in *G. lamblia* lysate (Figure 4). Using immunoprecipitation followed by

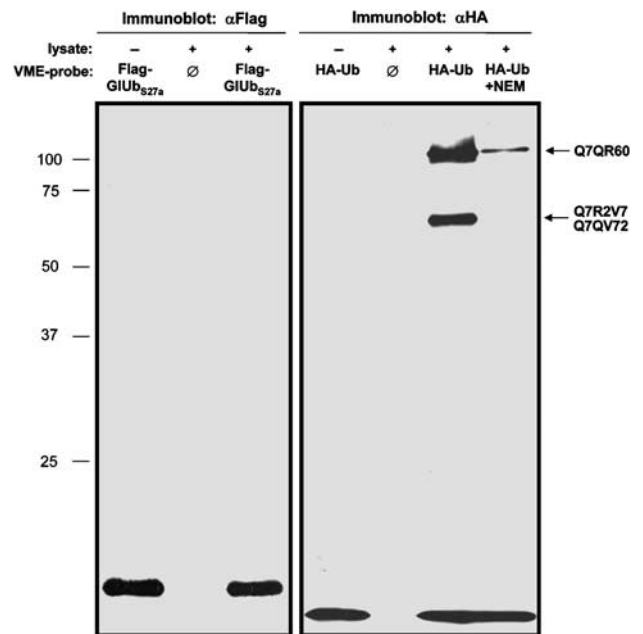


Figure 4 Evidence for ubiquitin-specific but not for GlUb_{S27a}-specific proteases in *G. lamblia*. 20 µg of *G. lamblia* lysate was incubated with 0.2 µg of electrophilic probes, in which HA-ubiquitin or 3 × Flag-GlUb_{S27a} were synthesized with a C-terminal vinylmethyl-ester trap (VME) (Borodovsky *et al*, 2002). Such traps react with substrate-specific cysteine proteases. The GlUb_{S27a}-based probe does not react with any proteins (left panel), whereas two dominant bands can be detected with a probe based on ubiquitin (right panel). Simultaneous addition of the cysteine-alkylating agent *N*-ethylmaleimide (5 mM) competes for binding to the ubiquitin trap. By immunoprecipitation and tandem-mass spectroscopy, the three polypeptides that react with ubiquitin-VME were identified as the putative deubiquitinating proteases Q7QR60 (93.1 kDa), Q7R2V7 (50.5 kDa), and Q7QV72 (49.9 kDa) (not shown). Samples were resolved by reducing SDS-PAGE (10%).

tandem mass spectroscopy, we identified the predominant proteins captured by this method. They represent three of the predicted seven *G. lamblia* deubiquitinases (Q7QR60, Q7R2V7, and Q7QV72, identified with 29.2, 27.2, and 5% sequence coverage, respectively). This hit rate agrees remarkably well with earlier experiments in yeast, mouse and human cells, where approximately 20–30% of the known ubiquitin-specific proteases react with such electrophilic ubiquitin-based probes (Borodovsky *et al*, 2002). In contradistinction, not a single of the putative *G. lamblia* deubiquitinases was labeled with the GIUb_{S27a}-based electrophile (Figure 4). Although we have no positive control for the reactivity of GIUb_{S27a}-VME, all other ubiquitin-like modifiers converted into electrophiles have yielded robust hits. Combined, our data are consistent with GIUb_{S27a} being a ubiquitin-like domain of S27a that is part of the holoribosome but not further processed by regulated proteolysis.

Discussion

The polyubiquitin gene encodes what is probably the most highly conserved protein in eukaryotes. Two other ubiquitin genes, encoding fusions to the ribosomal proteins S27a and L40, are less conserved, and variants of ubiquitin have been identified in protozoa, algae, amoeba, fungi, and in nematodes (Catic and Ploegh, 2005). We have analyzed the N-terminal domain of S27a in the ancient protist *G. lamblia*, featuring a protein domain that bears little sequence similarity with ubiquitin. Nonetheless, this protein moiety has completely conserved the ubiquitin fold, as determined by NMR spectroscopy. Other examples of proteins with this fold are the ubiquitin-like modifiers, which are covalently conjugated to target proteins in an ATP-dependent manner, and the ubiquitin-like domains, mostly found at the termini of large multidomain proteins (Buchberger, 2002). Ubiquitin-like modifiers share their genetic origin with ubiquitin, whereas some ubiquitin-like domains—for instance at the N-terminus of RAD23 or in the UBX protein family—might represent structural analogs rather than genetic homologs of ubiquitin (Buchberger, 2002; Iyer *et al*, 2006; Kiel and Serrano, 2006). In fact, the ubiquitin fold is one of nine originally defined superfold families, found in a diverse set of proteins, which do not necessarily share common functions (Orengo *et al*, 1994). The ubiquitin variant GIUb_{S27}, and with it possibly also other ubiquitin variants, embodies a third group: it is a descendant of ubiquitin, revealed by the conserved ribosomal S27a protein at its C-terminus, yet it does not participate in post-translational modifications. Moreover, its primary structure has deteriorated to the extent that its genetic origin is now concealed (Supplementary Figure 5). While the 3D structure of GIUb_{S27a} is equally compact and folds similarly as other ubiquitin-like domains, it is less complex in composition, consisting of fewer elements of ordered secondary structure (Supplementary Figure 6).

We have previously suggested that the polyubiquitin gene might have evolved as a concatemer to increase its recombinogenic potential (Catic and Ploegh, 2005). Paralogs of this gene have given rise to ubiquitin-like modifiers that—like ubiquitin—are mostly expressed as precursors with C-terminal extensions. Regulated proteolysis of the tail is required for biochemical activation of these modifiers. Early in eukaryote

evolution, predating the existence of *G. lamblia*, insertions of ubiquitin occurred at the N-terminus of the ribosomal proteins S27a and L40. The archaean homologs of both proteins lack N-terminal domains, and such ubiquitin moieties are not essential for viability of *S. cerevisiae* (Finley *et al*, 1989). Crosslinking studies suggest that S27a is not part of the catalytic core of the holoribosome (Tolan and Traut, 1981; Yeh *et al*, 1986; Takahashi *et al*, 2002), but its exact position remains to be determined (Spahn *et al*, 2001). However, the existence of ubiquitin or ubiquitin variants fused to S27a and L40, along with functional studies conducted in budding yeast, suggest that these domains do play a role in facilitating ribosome biogenesis (Ecker *et al*, 1989; Finley *et al*, 1989). The possible role of N-terminal ubiquitin(-like) domains as cotranslational chaperones has been reemphasized in a recent study by Varshavsky's group (Graciet *et al*, 2006). The authors provide a plausible scenario that explains how the addition of a ubiquitin fold could have stabilized the translation of suboptimally structured proteins very early in evolution. This idea is supported by the mainly N-terminal location of these domains in prokaryotes and phages (Iyer *et al*, 2006). Thus, while protein degradation in bacteria is efficiently dealt with by the N-end rule pathway in the absence of ubiquitin-conjugation, the presence of ubiquitin-like domains in such life forms might in part reflect their primordial function as facilitators of protein folding (Graciet *et al*, 2006).

In eukaryotes, however, Nei *et al* (2000) have discovered that the ubiquitin genes are under strong evolutionary pressure and preserve their sequences by purifying selection. Indeed, every eukaryote has at least one highly conserved ubiquitin gene. But how does this model fit with the existence of ubiquitin variants? Knowing that GIUb_{S27a} originated from ubiquitin, its sequence deterioration is impressive and apart from premature stop codons resembles the extent of variability seen in pseudogenes. In fact, we would argue that this domain is under little selection to preserve its primary structure. What may explain this lack of conservation could be the fact that GIUb_{S27a} is not a posttranslational modifier and is actually not even separated from the ribosome (Figure 5). Perhaps it was a random mutation, or structural inaccessibility, that made the original ubiquitin at the N-terminus of S27a in *G. lamblia* resistant to processing by deubiquitinating proteases. A single nucleotide exchange can generate such a protease-resistant mutant form of ubiquitin (Hodgins *et al*, 1992) by replacing the glycine at position 75 with alanine. Once stuck to the ribosome, the usual interactions of ubiquitin with proteins it would otherwise have access to as a posttranslational modifier become irrelevant, and structural constraints alone shape the further evolution of this domain. Our bioinformatics, functional and structural observations all support such a model. In that sense, a simple variation of the electrophilic GIUb_{S27a} probe that includes a C-terminal di-glycine motif as found in ubiquitin did not lead to recognition by *G. lamblia* proteases in a whole cell lysate (data not shown). Even if the insertion of the penultimate alanine were the initiating event that led to the uncleavable fusion, it is apparently not the only difference that prevents the fusion protein from being recognized by proteases currently found in *G. lamblia*. What logically follows from these data is the exciting possibility that other ubiquitin-like domains—even those unrelated to ubiquitin based on sequence

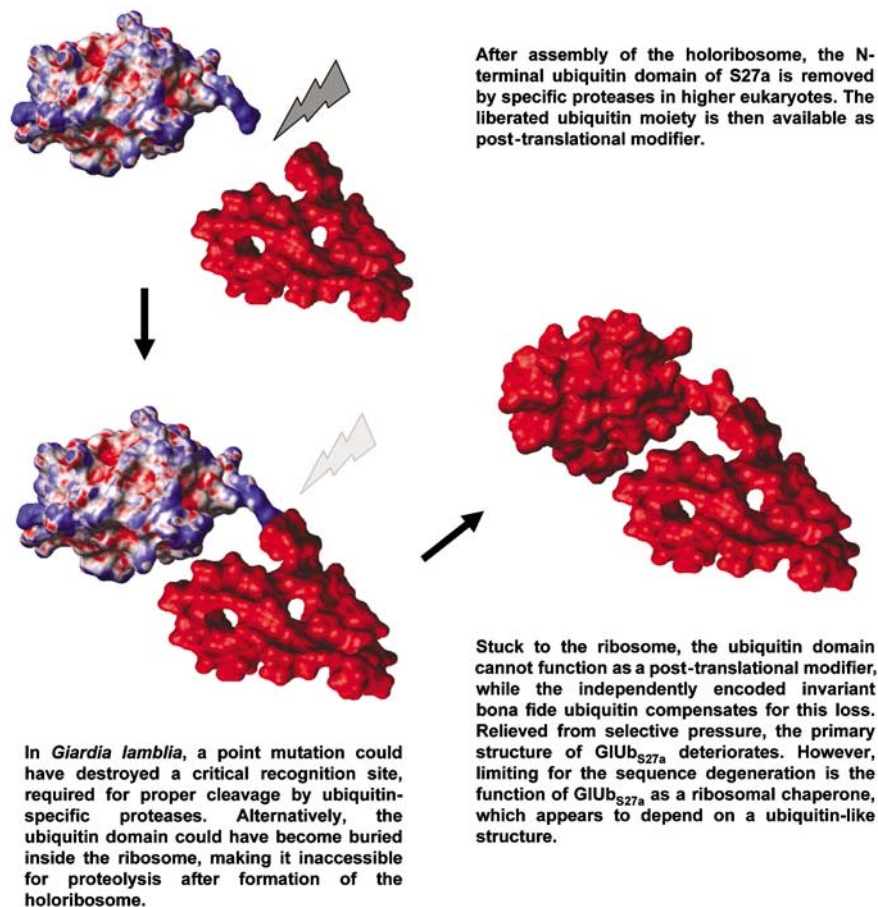


Figure 5 Model of the evolution of ubiquitin variants. We hypothesize that the irreversible fusion of ubiquitin to a larger protein exempts it from purifying selection and therefore allows for mutations at high rate. Ubiquitin is represented colored by surface charges; S27a and GIUb_{S27a} are colored in red. The structure of S27a has not been solved yet; we calculated a hypothetical model *ab initio* (Kim *et al*, 2004).

comparison alone—could have evolved by recombination of the polyubiquitin locus and the creation of a ubiquitin-fusion protein; if protease-resistant, it would have exempted the ubiquitin domain from purifying selection and allowed for sequence deterioration. Indeed, ancient ubiquitin-like domains, such as the UBX family, show little or no sequence similarity to ubiquitin, consistent with the absence of evolutionary pressure to preserve their primary structure. Furthermore, the ubiquitin-like domain of RAD23 can be functionally replaced with ubiquitin (Watkins *et al*, 1993), underscoring the importance of structure over sequence in this context. On the other hand, in more recently evolved ubiquitin-like domains, as for instance found at the tail of p59 OASL, acquired late during vertebrate evolution, the ubiquitin sequence fingerprint is still evident (Hartmann *et al*, 1998) (DOI 10.2210/pdb1wh3/pdb, to be published).

Taken together, we have shown that ubiquitin can exhibit remarkable plasticity in its primary structure, while maintaining its overall secondary and tertiary structure. Such sequence variation might depend on ubiquitin being fused to a larger protein in a protease-resistant manner, therefore rendering it immune to purifying selection. Considering the vast sequence deterioration of GIUb_{S27a}—undoubtedly a former bona fide ubiquitin—there is no reason to assume that other, by sequence unrelated, ubiquitin-like domains could not as well have evolved in a similar manner, rather than be independent structural analogs of ubiquitin. Hence, the

sweeping success of the ubiquitin superfold in eukaryotes might in part be a consequence of the polyubiquitin locus' propagation throughout the genome by recombination.

Materials and methods

Structure determination

The gene encoding *G. lamblia* Ub_{S27} (67 amino acids) was synthesized using *Escherichia coli*-optimized codons. Backtranslation was performed with the toolbox offered by www.entelechon.com. Oligonucleotides of up to 70 bases were synthesized with 34 or 35 base overlaps, 5'-phosphorylated (Integrated DNA Technologies) and ligated with T4 ligase (New England Biolabs). Protein samples for NMR experiments were overexpressed in the BL21 (DE3) strain of *E. coli* by using the pET28a plasmid (both Novagen), containing DNA sequences of two constructs encoding GIUb_{S27A}: we had cloned GIUb_{S27A} either with a 3 × Flag tag at its N-terminus, following the His tag of the pET28a vector, or with the GB1 domain at the N-terminus and the His tag at the very C-terminus of this fusion. The 3 × Flag-tagged construct was used in the preliminary secondary structure assignment, while the GB1-tagged construct was used for the complete structure determination. The GB1 domain derives from the IgG-binding protein in *Streptococcus* spp, is well defined by NMR and highly soluble (Zhou *et al*, 2001). The GIUb_{S27A} domain could not be expressed in *E. coli* when using only a His tag or a single HA tag. Furthermore, a 3 × Flag-GIUb_{S27A} construct, cloned into pcDNA3.1, could not be expressed in mammalian cells (verified by immunoblot and by immunohistochemistry).

Uniformly ¹⁵N/¹³C-labeled NMR samples were expressed in bacteria grown in minimal M9 media, with ¹⁵NH₄Cl and ¹³C-glucose

as sole nitrogen and carbon sources. The cells were cultivated at 36°C to an OD of 0.6 and then induced with 1 mM IPTG for 18 h. The GIUB_{S27A} proteins were isolated by lysing the bacteria in a French Press, recovering of the proteins with a Nickel-NTA column (Qiagen) and FPLC purification to >95% (assessed by Coomassie staining), as previously described (Misaghi *et al*, 2004). The NMR samples were prepared in 10 mM phosphate buffer, pH 6.2, 100 mM NaCl with a typical concentration of 1 mM GIUB_{S27A} and remained stable for about one week at 25°C.

NMR experiments for GB1-tagged GIUB_{S27A} include: HNCA, HN(CA)CB, HNCO for the backbone assignment in non-uniformly sampled manner (Rovnyak *et al*, 2004), C(CO)NH, H(CCO)NH, HCCH-TOCSY for the side-chain assignment, and ¹⁵N-edited NOESY, ¹³C-edited aliphatic and aromatic NOESY, as well as 2D-NOESY for obtaining distance constraints using conventional methodology (Ferentz and Wagner, 2000). These experiments were performed on a Varian Inova600 equipped with a chilly-probe at 25°C, with the exception of the ¹⁵N-NOESY and 2D-NOESY, which were performed on the Bruker750 spectrometer, equipped with a cryo-probe at 25°C.

NMR data sets were processed using the software PROSA (Güntert *et al*, 1992) for the conventional experiments, and RNMTK (<http://www.rowland.harvard.edu/rnmrtk/toolkit.html>) for the non-uniformly sampled experiments. Data were analyzed using the software CARA (<http://www.nmr.ch/>). The secondary structures were determined with the software TALOS (Cornilescu *et al*, 1999), using the chemical shift values of CA, CB, HA, and N. The structural model for GIUB_{S27A} was calculated with the software CYANA (Güntert *et al*, 1997) using 84 TALOS angular constraints, and 830 NOE constraints assigned by CARA (Table II). Structure models were visualized with MolMol (<http://hugin.ethz.ch/wuthrich/software/molmol/>) (Koradi *et al*, 1996).

Immunoblots

G. lamblia trophozoites (strain MR4) were cultured axenically under standard conditions (Ghosh *et al*, 2001), harvested, washed

3 × with ice-cold PBS, flash frozen and stored at -80°C. Cellular lysate was obtained by resuspending trophozoites in lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 2 mM DTT) and intermittent vortexing in the presence of glassbeads and 0.5% reduced Triton X-100 (both Sigma-Aldrich) for 30 min. All biochemical procedures were performed on ice or at 4°C, and protein content was determined with the BCA kit (Pierce). Functional polyclonal anti-GIUB_{S27A} serum was obtained by inoculating a rabbit with 1 mg of GB1-GIUB_{S27A} according to standard protocol (Covance, study nr. 0053523, animal 5948). Immunoblots were conducted using a 1:100 dilution of the serum in PBST with 0.5% milk overnight at 4°C, followed by 60 min incubation at room temperature with an HRP-conjugated anti-rabbit IgG serum (1:20 000) (Southern Biotech). As competing antigen, we used 3 × Flag-GIUB_{S27A} without a His tag (generated by hydrolysis of the intein-precursor to the electrophilic probe, see below), to exclude cross-reactivity of the rabbit serum with either the GB1 domain or the His tag, both present in the initially injected antigen. As a molecular weight standard, we used the same batch of *Precision Plus Dual Color* (Biorad) throughout all experiments.

Ribosome isolation

A 40 ml volume of densely grown MR4 trophozoites was harvested, washed, and immediately lysed with glassbeads to obtain a total of 5 mg protein in 500 µl. The lysis buffer contained 1 M NH₄Ac, 20 mM Tris, pH 7.6, 5 mM *N*-ethylmaleimide to inhibit endosomal cysteine proteases, and 15 mM MgAc to stabilize the holoribosome (dissociation of the main RNA peak in the absence of MgAc was used to control for the isolation procedure). For the fractionation, 14 ml 10–40% sucrose gradient were centrifuged for 2 h (288 000 *g*_{max}, SW41 Ti swinging-bucket), as previously described (Shirakura *et al*, 2001). Equal fractions of 250 µl were isolated with a gradient fractionator (Biocomp, Ver. 3.0) and analyzed by UV spectroscopy (OD 260 nm). OD 330 nm spectra were used to correct for light scattering.

Probe generation

3 × Flag-GIUB_{S27A} was cloned into pTYB2 (New England Biolabs) in the absence of a His tag. The C-terminal glycine G67 was removed by site-directed mutagenesis (Phusion SDM kit, New England Biolabs) and the probe was generated using glycine-vinylmethyl-ester, as previously described (Borodovsky *et al*, 2002). The MESNa adduct was obtained by incubation overnight at 37°C and conversion with the reactive warhead was achieved after 4–5 h incubation at 37°C. Progress was monitored by size shift in 15% SDS-PAGE after insertion of glycine-VME. Labeling of cell lysate (see immunoblot section above for lysis buffer) with the probe was performed on ice for 30 min. No additional bands were visible when incubating at higher temperatures, but the stability of the lysate was greatly reduced at room temperature or at 37°C, due to the high amount of endosomal proteases in *G. lamblia*. Immunoblots were performed using anti-Flag M2 antibodies (Sigma) or anti-HA clone 12CA5. For immunoprecipitation, 100 µg lysate were incubated with 5 µg HA-ubiquitin-VME for 30 min on ice. The mixture was rocked for 4 h at 4°C in the presence of equilibrated anti-HA beads (Roche, 3F10-agarose, 50 µl slurry). The precipitate was washed three times with NET buffer (50 mM Tris, pH 7.4, 0.5% NP40, 150 mM NaCl, 5 mM EDTA) and eluted by boiling in SDS sample buffer. After SDS-PAGE, gel staining, and protein trypsinolysis (Borodovsky *et al*, 2002), the recovered peptides were analyzed by reverse-phase liquid chromatography electrospray ionization mass spectrometry using a Waters NANO-ACQUITY-UPLC, coupled to a Thermo LTQ linear ion-trap mass spectrometer. MS/MS spectra were searched against the non-redundant NCBI database using SEQUEST (<http://fields.scripps.edu/quest/>). SEQUEST results were analyzed with Bioworks Browser 3.2 and filtered with the following criteria: different peptides, minimum cross correlation coefficients of 1.50, 2.00, and 2.50 at 1, 2, and 3 charge states, respectively, and a minimum of two different peptides per protein with an Sp preliminary score of 300 or higher.

Bioinformatics

Protein sequences were obtained from the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) and the *G. lamblia* Genome Database (www.mbl.edu/Giardia) (McArthur

Table II NMR and refinement statistics for protein structures

	GIUB _{S27A}
<i>NMR distance and dihedral constraints</i>	
<i>Distance constraints</i>	
Total NOE	830
Intra-residue	487
Inter-residue	343
Sequential ($ i-j = 1$)	191
Medium-range ($ i-j < 4$)	51
Long-range ($ i-j > 5$)	101
Hydrogen bonds	24
<i>Total dihedral angle restraints</i>	84
phi	42
psi	42
<i>Ramachandran statistics</i>	
Most favored region (%)	80.7
Additionally allowed region (%)	18.3
Generously allowed region (%)	0.8
Disallowed region (%)	0.2
<i>Structure statistics^a</i>	
<i>Violations</i>	
Max. dihedral angle violation (deg)	< 5
Max. distance constraint violation (Å)	< 0.5
<i>Deviations from idealized geometry</i>	
Bond lengths (Å)	0.001
Bond angles (deg)	0.2
<i>Average pairwise r.m.s.d. (mean and s.d., in Å)</i>	
Heavy	1.41 ± 0.13
Backbone	0.76 ± 0.16

^aTen lowest energy structures selected from 50 calculated by CYANA.

et al, 2000). Sequence names are given in SWISSPROT designation or—if not available—in the ORF number according to the *G. lamblia* Genome Database. Protein sequences were aligned with ClustalX (<http://bips.u-strasbg.fr/fr/Documentation/ClustalX>) (Thompson *et al*, 1997), and visualized with JalView (www.jalview.org) (Clamp *et al*, 2004). The UPGMA phylogram was calculated with the MEGA 3.1 software package (www.megasoftware.net) with 100 bootstrap replicates, pairwise deletion and homogeneous substitution according to the JTT model (Kumar *et al*, 2004). This setting was also used for creation of the pairwise distance matrix. The Maximum Likelihood phylogram was generated using PhyML (<http://atgc.lirmm.fr/phyml/>) (Guindon and Gascuel, 2003), following bootstrapping by SEQBOOT from the PhyIP package (<http://evolution.genetics.washington.edu/phyip.html>) (Felsenstein, 1988). Settings were used as previously established for ubiquitin-like modifiers (Xu *et al*, 2006). The nearest neighbor analysis for ubiquitin conjugating enzymes was carried out with the N-J tree option of ClustalX.

References

Archibald JM, Teh EM, Keeling PJ (2003) Novel ubiquitin fusion proteins: ribosomal protein P1 and actin. *J Mol Biol* **328**: 771–778

Borodovsky A, Ovaas H, Kollig N, Gan-Erdene T, Wilkinson KD, Ploegh HL, Kessler BM (2002) Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem Biol* **9**: 1149–1159

Buchberger A (2002) From UBA to UBX: new words in the ubiquitin vocabulary. *Trends Cell Biol* **12**: 216–221

Catic A, Ploegh HL (2005) Ubiquitin—conserved protein or selfish gene? *Trends Biochem Sci* **30**: 600–604

Chan YL, Suzuki K, Wool IG (1995) The carboxyl extensions of two rat ubiquitin fusion proteins are ribosomal proteins S27a and L40. *Biochem Biophys Res Commun* **215**: 682–690

Clamp M, Cuff J, Searle SM, Barton GJ (2004) The Jalview Java alignment editor. *Bioinformatics* **20**: 426–427

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**: 289–302

Ecker DJ, Stadel JM, Butt TR, Marsh JA, Monia BP, Powers DA, Gorman JA, Clark PE, Warren F, Shatzman A, Crooke ST (1989) Increasing gene expression in yeast by fusion to ubiquitin. *J Biol Chem* **264**: 7715–7719

Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet* **22**: 521–565

Ferentz AE, Wagner G (2000) NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q Rev Biophys* **33**: 29–65

Finley D, Bartel B, Varshavsky A (1989) The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. *Nature* **338**: 394–401

Gallego E, Alvarado M, Wasserman M (2007) Identification and expression of the protein ubiquitination system in *Giardia intestinalis*. *Parasitol Res* **101**: 1–7

Ghosh S, Frisardi M, Rogers R, Samuelson J (2001) How *Giardia* swim and divide. *Infect Immun* **69**: 7866–7872

Graciet E, Hu RG, Piatkov K, Rhee JH, Schwarz EM, Varshavsky A (2006) Aminoacyl-transferases and the N-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen. *Proc Natl Acad Sci USA* **103**: 3078–3083

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704

Güntert P, Doetsch V, Wider G, Wüthrich K (1992) Processing of multi-dimensional NMR data with the new software PROSA. *J Biomol NMR* **2**: 619–629

Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**: 283–298

Hartmann R, Olsen HS, Widder S, Jorgensen R, Justesen J (1998) p50ASL, a 2'-5' oligoadenylate synthetase like protein: a novel human gene related to the 2'-5' oligoadenylate synthetase family. *Nucleic Acids Res* **26**: 4121–4128

Hodgins RR, Ellison KS, Ellison MJ (1992) Expression of a ubiquitin derivative that conjugates to protein irreversibly produces

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

We thank Sarah Ross from the Gillin lab (UC San Diego) for providing us with *G. lamblia* trophozoites, and appreciate comments from Paul Galardy (Mayo Clinic), Greg Korbel, Britta Müller, Victor Quesada, and other members of the Ploegh lab. We are grateful to Sean Moore from Robert Sauer's laboratory (MIT) for technical support in ribosome purification. Finally, we thank Alexander Varshavsky (Caltech) and Daniel Finley (Harvard Medical School) for helpful suggestions. This work was supported by NIH grants to HLP, to GW (CA68262 and GM47467), to JS (AI48082), and by the Training Program in Host Pathogen Interaction to DMR (5T32 AI052070).

phenotypes consistent with a ubiquitin deficiency. *J Biol Chem* **267**: 8807–8812

Iyer LM, Burroughs AM, Aravind L (2006) The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like beta-grasp domains. *Genome Biol* **7**: R60

Jentsch S, Pyrowolakis G (2000) Ubiquitin and its kin: how close are the family ties? *Trends Cell Biol* **10**: 335–342

Jones D, Candido EP (1993) Novel ubiquitin-like ribosomal protein fusion genes from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *J Biol Chem* **268**: 19545–19551

Kerscher O, Felberbaum R, Hochstrasser M (2006) Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol* **22**: 159–180

Kiel C, Serrano L (2006) The ubiquitin domain superfold: structure-based sequence alignments and characterization of binding epitopes. *J Mol Biol* **355**: 821–844

Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32**: W526–W531

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**: 29–32, 51–55

Krebber H, Wostmann C, Bakker-Grunwald T (1994) Evidence for the existence of a single ubiquitin gene in *Giardia lamblia*. *FEBS Lett* **343**: 234–236

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**: 150–163

Lake MW, Wuebbens MM, Rajagopalan KV, Schindelin H (2001) Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. *Nature* **414**: 325–329

Louie DF, Resing KA, Lewis TS, Ahn NG (1996) Mass spectrometric analysis of 40 S ribosomal proteins from Rat-1 fibroblasts. *J Biol Chem* **271**: 28189–28198

McArthur AG, Morrison HG, Nixon JE, Passamaneck NQ, Kim U, Hinkle G, Crocker MK, Holder ME, Farr R, Reich CI, Olsen GE, Aley SB, Adam RD, Gillin FD, Sogin ML (2000) The *Giardia* genome project database. *FEMS Microbiol Lett* **189**: 271–273

Misaghi S, Sun ZY, Stern P, Gaudet R, Wagner G, Ploegh H (2004) Structural and functional analysis of human cytomegalovirus US3 protein. *J Virol* **78**: 413–423

Nei M, Rogozin IB, Piontkivska H (2000) Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci USA* **97**: 10866–10871

Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* **39**: 121–152

Orengo CA, Jones DT, Thornton JM (1994) Protein superfamilies and domain superfolds. *Nature* **372**: 631–634

Ramage R, Green J, Muir TW, Ogunjobi OM, Love S, Shaw K (1994) Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem J* **299** (Part 1): 151–158

Redman KL, Rechsteiner M (1989) Identification of the long ubiquitin extension as ribosomal protein S27a. *Nature* **338**: 438–440

- Rovnyak D, Frueh DP, Sastry M, Sun ZY, Stern AS, Hoch JC, Wagner G (2004) Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *J Magn Reson* **170**: 15–21
- Shirakura T, Maki Y, Yoshida H, Arisue N, Wada A, Sanchez LB, Nakamura F, Muller M, Hashimoto T (2001) Characterization of the ribosomal proteins of the amitochondriate protist, *Giardia lamblia*. *Mol Biochem Parasitol* **112**: 153–156
- Sloper-Mould KE, Jemc JC, Pickart CM, Hicke L (2001) Distinct functional surface regions on ubiquitin. *J Biol Chem* **276**: 30483–30489
- Spahn CM, Beckmann R, Eswar N, Penczek PA, Sali A, Blobel G, Frank J (2001) Structure of the 80S ribosome from *Saccharomyces cerevisiae*—tRNA-ribosome and subunit-subunit interactions. *Cell* **107**: 373–386
- Spence J, Gali RR, Dittmar G, Sherman F, Karin M, Finley D (2000) Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell* **102**: 67–76
- Takahashi Y, Mitsuma T, Hirayama S, Odani S (2002) Identification of the ribosomal proteins present in the vicinity of globin mRNA in the 40S initiation complex. *J Biochem (Tokyo)* **132**: 705–711
- Tatusova TA, Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* **174**: 247–250
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882
- Tolan DR, Traut RR (1981) Protein topography of the 40 S ribosomal subunit from rabbit reticulocytes shown by cross-linking with 2-iminothiolane. *J Biol Chem* **256**: 10129–10136
- Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol* **194**: 531–544
- Vladimirov SN, Ivanov AV, Karpova GG, Musolyamov AK, Egorov TA, Thiede B, Wittmann-Liebold B, Otto A (1996) Characterization of the human small-ribosomal-subunit proteins by N-terminal and internal sequencing, and mass spectrometry. *Eur J Biochem* **239**: 144–149
- Watkins JF, Sung P, Prakash L, Prakash S (1993) The *Saccharomyces cerevisiae* DNA repair gene RAD23 encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Mol Cell Biol* **13**: 7757–7765
- Went HM, Jackson SE (2005) Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel* **18**: 229–237
- Xu J, Zhang J, Wang L, Zhou J, Huang H, Wu J, Zhong Y, Shi Y (2006) Solution structure of Urm1 and its implications for the origin of protein modifiers. *Proc Natl Acad Sci USA* **103**: 11625–11630
- Yeh YC, Traut RR, Lee JC (1986) Protein topography of the 40 S ribosomal subunit from *Saccharomyces cerevisiae* as shown by chemical cross-linking. *J Biol Chem* **261**: 14148–14153
- Zhou P, Lugovskoy AA, Wagner G (2001) A solubility enhancement tag (SET) for NMR studies of poorly behaving proteins. *J Biomol NMR* **20**: 11–14